PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA
MINISTRY OF HIGHER EDUCATION
AND SCIENTIFIC RESEARCH
University of Echahid Hamma Lakhdar - El Oued

Department of process engineering and petrochemicals

MASTER'S THESIS

Submitted in partial fulfilment of the requirements for the award of the Academic Master's Degree (LMD).

Startup Project

Submitted in partial fulfilment of the requirements for obtaining the Startup Creation Project Certificate under Ministerial Decision No. 1275.

Domain: Science and Technology

Division: Process Engineering

Specialty: Chemical Engineering

Presented by:

ZEGHIB Mohammed Assil

Entitled:

Spectra AI: Development of an Intelligent Assistant for Automated Interpretation of FTIR Spectra and AI-Based Generation of Scientific Reports

Publicly defended in: 06/06/2026
The jury composed of:

REDJEB Youcef	MCA	President
FERHAT Mohammed Fouad	Full Professor	Examiner
TEDJANI Mohammed Laid	MCB	Supervisor
DJEDIDI Mohammed Reda	/	Economic partner representative
BAHI Abdelmalek	MCB	University interfaces representative

Academic Year: 2025/2026


بسم الله الرحمن الرحيم

Dedication

To my beloved parents, the source of my strength and inspiration,

To my older and little sisters, my lifelong friends and supporters,

To all the members of my extended family,

To the cherished memory of my grandparents,

To all my esteemed professors and instructors,

To my friends and classmates who shared this journey,

To the relentless pursuit of knowledge, the thrill of scientific discovery,

and the hope of applying it for the betterment of humankind.

Acknowledgments

First and foremost, I praise Allah, the Almighty, for giving me the strength, patience, and health to complete this work.

I would like to express my deepest gratitude to my supervisor, ***Prof. TEDJANI Mohammed Laid****, for his invaluable guidance, constant support, and continuous encouragement throughout this research. His profound knowledge, insightful advice, and constructive feedback have been instrumental in shaping this thesis, and it has been an absolute privilege to work under his supervision.*

I am sincerely grateful to the members of the jury for accepting to review and evaluate this work, and for their valuable time and constructive remarks.

My appreciation also goes to all the professors and instructors in the Department of Process Engineering at Echahid Hamma Lakhdar University for their high-quality education and dedication.

I would also like to extend my heartfelt thanks to all my friends and classmates who shared this academic journey with me. A very special thanks goes to my dear friend, ***MOHAMMED MAHFOUDH BEDDADA****, for his unwavering support, motivating presence, and for always being a source of encouragement.*

Finally, words cannot express my profound gratitude and utmost respect to my beloved parents. Their unconditional love, endless sacrifices, and continuous prayers have been the foundation of my life and the driving force behind every success I achieve.

Abstract

Fourier Transform Infrared spectroscopy is among the most widely employed analytical techniques in materials science and chemical engineering, yet the workflow connecting raw spectral data to publication-ready scientific interpretation remains largely manual, time-consuming, and dependent on specialist expertise. This thesis presents the design, implementation, and experimental validation of SPECTRA AI, an intelligent desktop assistant developed in Python that automates the principal steps of FTIR spectral analysis within a single accessible graphical user interface. The system integrates a prominence-based automated peak detection algorithm implemented through the SciPy signal processing library, an interactive spectral visualisation engine built on Matplotlib, and an AI-powered scientific report generation module connected to the Google Gemini large language model via REST API. A structured prompt engineering methodology was developed to elicit journal-conformant Results and Discussion text following the stylistic conventions of Elsevier materials science publications, with IUPAC vibrational nomenclature and citation placeholder integration. The tool was validated against published peer-reviewed FTIR spectra of three nanomaterial systems of contrasting spectral complexity: biosynthesised Ag/AgCl nanoparticles, CuO@Tocopherol hybrid nanoparticles, and ZnO/$BaMg_2$ mixed-oxide nanoparticles. Validation results demonstrated a 100% recall of all labelled reference absorption bands across the three case studies, with an overall mean positional error below 15 cm^{-1} and sub-resolution accuracy of 3.7 cm^{-1} on sharp, well-defined features. The system exports results in multiple formats including plain text, Excel spreadsheet, and formatted Microsoft Word document embedding the annotated spectrum, the AI-generated discussion, and a structured peak assignment table. A business model and operational plan for commercial deployment of the tool are presented as an annex. The work demonstrates that the integration of established signal processing algorithms with modern generative artificial intelligence constitutes a viable, practically valuable, and deployable approach to reducing the analytical burden of routine FTIR characterisation in academic research and industrial quality control environments.

Keywords: FTIR spectroscopy, peak detection, artificial intelligence, large language model, scientific report generation, nanomaterials, Python, Gemini API, prompt engineering, spectral analysis automation.

Résumé

La spectroscopie infrarouge à transformée de Fourier figure parmi les techniques analytiques les plus utilisées en science des matériaux et en génie chimique. Pourtant, le processus reliant les données spectrales brutes à une interprétation scientifique prête à la publication demeure largement manuel, chronophage et tributaire d'une expertise spécialisée. Ce mémoire présente la conception, la mise en œuvre et la validation expérimentale de SPECTRA AI, un assistant de bureau intelligent développé en Python qui automatise les principales étapes de l'analyse spectrale FTIR au sein d'une interface graphique unique et accessible. Le système intègre un algorithme automatisé de détection de pics basé sur la proéminence, implémenté via la bibliothèque de traitement du signal SciPy, un moteur de visualisation spectrale interactif construit sur Matplotlib, ainsi qu'un module de génération de rapports scientifiques alimenté par le modèle de langage Google Gemini, accessible via une API REST. Une méthodologie d'ingénierie de prompt structurée a été développée afin d'obtenir des textes de type Résultats et Discussion conformes aux conventions stylistiques des revues Elsevier en science des matériaux, intégrant la nomenclature vibrationnelle IUPAC et des espaces réservés aux références bibliographiques. L'outil a été validé sur des spectres FTIR publiés dans des revues à comité de lecture, portant sur trois systèmes de nanomatériaux de complexité spectrale contrastée : des nanoparticules d'Ag/AgCl biosynthétisées, des nanoparticules hybrides CuO@Tocophérol, et des nanoparticules d'oxydes mixtes ZnO/$BaMg_2$. Les résultats de validation ont démontré un taux de rappel de 100 % pour toutes les bandes d'absorption de référence étiquetées dans les trois études de cas, avec une erreur positionnelle moyenne globale inférieure à 15 cm^{-1} et une précision sous-résolution de 3,7 cm^{-1} pour les bandes nettes et bien définies. Le système exporte les résultats en plusieurs formats, notamment en texte brut, en feuille de calcul Excel et en document Microsoft Word formaté incorporant le spectre annoté, la discussion générée par l'IA et un tableau structuré d'attribution des pics. Un modèle économique et un plan opérationnel pour le déploiement commercial de l'outil sont présentés en annexe. Ce travail démontre que l'intégration d'algorithmes de traitement du signal éprouvés avec l'intelligence artificielle générative moderne constitue une approche viable, pratiquement utile et déployable pour réduire la charge analytique de la caractérisation FTIR de routine dans les environnements de recherche académique et de contrôle qualité industriel.

Mots-clés : spectroscopie FTIR, détection de pics, intelligence artificielle, modèle de langage, génération de rapports scientifiques, nanomatériaux, Python, API Gemini, ingénierie de prompt, automatisation de l'analyse spectrale.

ملخص

تُعدّ مطيافية الأشعة تحت الحمراء بتحويل فورييه (FTIR) من أكثر التقنيات التحليلية استخدامًا في علوم المواد والهندسة الكيميائية. ومع ذلك، فإن سير العمل الذي يربط البيانات الطيفية الخام بالتفسير العلمي الجاهز للنشر لا يزال يُنجز بصورة يدوية إلى حدٍّ كبير، مما يجعله مستهلكًا للوقت ومعتمدًا بدرجة كبيرة على الخبرة المتخصصة. يقدّم هذا العمل تصميمًا وتطويرًا وتحقيقًا تجريبيًا لنظام **SPECTRA AI**، وهو مساعد ذكي لسطح المكتب مطوَّر بلغة Python يهدف إلى أتمتة المراحل الرئيسة لتحليل أطياف FTIR ضمن واجهة رسومية موحّدة وسهلة الاستخدام.

يدمج النظام خوارزمية آلية للكشف عن القمم الطيفية تعتمد على مفهوم البروزية(Prominence) ، وقد نُفِّذت باستخدام مكتبة معالجة الإشاراتSciPy ، إلى جانب محرك تفاعلي لعرض الأطياف وتفسيرها مبني علىMatplotlib ، ووحدة لتوليد التقارير العلمية مدعومة بنموذج اللغة الكبير Google Gemini عبر واجهة برمجة تطبيقات .REST كما جرى تطوير منهجية منظّمة لهندسة الأوامر النصية بهدف إنتاج نصوص علمية من نوع «النتائج والمناقشة» تتوافق مع الأعراف الأسلوبية المتبعة في مجلات Elsevier المتخصصة في علوم المواد، مع دمج التعيين الاهتزازي للمجاميع الوظيفية وفق توصيات IUPAC وإدراج مواضع مرجعية مخصّصة للاستشهادات الببليوغرافية.

تم التحقق من أداء النظام من خلال مقارنة نتائجه بأطياف FTIR منشورة في مجلات علمية محكّمة لثلاثة أنظمة من الجسيمات النانوية ذات مستويات مختلفة من التعقيد الطيفي، وهي: جسيمات Ag/AgCl النانوية المُحضَّرة حيويًا، وجسيمات CuO@Tocopherol الهجينة، وجسيمات $ZnO/BaMg_2$ ذات الأكاسيد المختلطة. وأظهرت نتائج التحقق معدل استرجاع بلغ 100% لجميع نطاقات الامتصاص المرجعية المعلَّمة في الدراسات الثلاث، مع متوسط خطأ موضعي إجمالي أقل من 15 سم$^{-1}$، ودقة بلغت 3.7 سم$^{-1}$ للنطاقات الحادة والمحددة جيدًا، وهي قيمة أقل من دقة العديد من الأجهزة المستخدمة في القياس.

يتيح النظام تصدير النتائج بصيغ متعددة تشمل النصوص العادية، وجداولMicrosoft Excel ، ووثائق Microsoft Wordالمنسّقة التي تتضمن الطيف المشروح، والمناقشة المولَّدة بالذكاء الاصطناعي، وجدولًا منظّمًا للقمم الطيفية وتفسيراتها. كما يُقدَّم في ملحق مستقل نموذج أعمال وخطة تشغيلية لاستثمار الأداة تجاريًا.

تُثبت نتائج هذا العمل أن دمج خوارزميات معالجة الإشارات الراسخة مع تقنيات الذكاء الاصطناعي التوليدي الحديثة يمثّل نهجًا عمليًا وقابلًا للتطبيق والنشر الفعلي، ويسهم في تقليل العبء التحليلي المرتبط بالتوصيف الروتيني بأطياف FTIR في بيئات البحث الأكاديمي ومختبرات ضبط الجودة الصناعية.

الكلمات المفتاحية :مطيافية الأشعة تحت الحمراء بتحويل فورييه(FTIR) ، كشف القمم الطيفية، الذكاء الاصطناعي، نماذج اللغة الكبيرة، توليد التقارير العلمية، المواد النانوية، Python، Google Gemini، هندسة الأوامر النصية، أتمتة التحليل الطيفي.

Table of Contents

List of Abbreviations

Abbreviation	Full Form
AI	Artificial Intelligence
API	Application Programming Interface
ATR	Attenuated Total Reflectance
CSV	Comma-Separated Values
EDX	Energy-Dispersive X-ray Spectroscopy
FFT	Fast Fourier Transform
FIR	Far Infrared
FTIR	Fourier Transform Infrared
FWHM	Full Width at Half Maximum
GUI	Graphical User Interface
HTTP	Hypertext Transfer Protocol
IUPAC	International Union of Pure and Applied Chemistry
KBr	Potassium Bromide
LLM	Large Language Model
LMD	Licence-Master-Doctorat (Bologna Process system)
MIR	Mid Infrared
NaCl	Sodium Chloride
NIR	Near Infrared
NP	Nanoparticle
OOXML	Office Open XML
PNG	Portable Network Graphics
REST	Representational State Transfer
SEM	Scanning Electron Microscopy
TGA	Thermogravimetric Analysis
UV-Vis	Ultraviolet-Visible Spectroscopy
XRD	X-ray Diffraction
ZnSe	Zinc Selenide

List of Figures

List of tables

General Introduction

General Introduction

Infrared spectroscopy, and Fourier Transform Infrared spectroscopy in particular, occupies a central position in the analytical toolkit of modern materials science and chemical engineering. Its capacity to provide a molecular fingerprint of virtually any solid, liquid, or gaseous samplerapidly, non-destructively, and with minimal sample preparationhas made it one of the most routinely performed characterisation techniques in research laboratories worldwide. From the synthesis of metal oxide nanoparticles to the functionalisation of biopolymers, from the development of composite nanomaterials to the confirmation of surface coatings, FTIR spectroscopy serves as an indispensable first line of structural verification, providing direct evidence of chemical bonding, functional group identity, phase formation, and interfacial interactions.

Despite its technical maturity, the analytical workflow surrounding FTIR spectroscopy has remained surprisingly labour-intensive. Once an instrument delivers a raw spectrum, the researcher is expected to manually identify absorption peaks, assign each feature to its corresponding vibrational mode through consultation of reference tables and the primary literature, evaluate the chemical significance of the observed pattern, and ultimately translate these observations into structured scientific prose suitable for journal publication. This process demands both spectroscopic expertise and considerable timetwo resources that are not always available in equal measure, particularly for graduate students and early-stage researchers working under the dual pressure of experimental productivity and manuscript preparation deadlines.

The emergence of artificial intelligence, and specifically of large language models capable of generating coherent, contextually accurate scientific text, opens a genuinely new possibility: that the interpretive and communicative steps of spectral analysis could be meaningfully automated without sacrificing scientific rigour. If a computational system could reliably detect the absorption peaks of an FTIR spectrum, assign them to their functional group origins, and produce a draft Results and Discussion section conforming to the conventions of peer-reviewed publication, it would represent a significant reduction in the analytical burden carried by researchers across chemistry, materials science, and chemical engineering. Furthermore, if such a system were conceived not merely as an academic prototype but as a deployable, commercially viable product, it could reach a far broader community of beneficiaries spanning

university laboratories, pharmaceutical quality control, industrial process monitoring, and environmental analysis.

In spite of the commercial availability of FTIR instruments and their widespread deployment in academic and industrial laboratories, no open-access integrated software tool currently exists that combines automated spectral peak detection with AI-driven scientific report generation within a single desktop environment accessible to non-specialist users. Commercial spectral analysis packages such as PerkinElmer Spectrum and Bruker OPUS provide powerful data processing capabilities but offer no pathway from spectral data to written scientific interpretation. Conversely, general-purpose large language models, when queried directly about FTIR spectra, lack the structured numerical input and domain-specific prompt engineering necessary to produce reliable, peak-specific scientific text. The gap between raw instrument output and manuscript-ready interpretation therefore remains openand it is precisely this gap, both technical and commercial, that the present work seeks to address.

Objectives of the Work

The principal objective of this thesis is the design, implementation, experimental validation, and commercial conceptualisation of an intelligent desktop assistantdesignated SPECTRA AIcapable of performing the complete FTIR analytical workflow from raw spectral data ingestion to the generation of publication-ready scientific reports. This overarching objective is pursued through five specific aims:

1. To establish the theoretical and instrumental foundations of FTIR spectroscopy as the scientific basis upon which the tool is designed, covering infrared absorption principles, Fourier transform instrumentation, and standard spectral interpretation conventions.
2. To implement a robust, parameter-adjustable automated peak detection algorithm capable of reliably identifying absorption bands across the full mid-infrared range for diverse material classes including metal oxide nanoparticles, biopolymers, carbon nanomaterials, and hybrid composite systems, integrated within an accessible graphical user interface.
3. To develop and optimise a prompt engineering methodology that enables a large language modelspecifically the Google Gemini APIto generate structured, journal-conformant FTIR Results and Discussion text from numerical peak data, following the stylistic and citation conventions of Elsevier materials science journals.

4. To validate the performance of the implemented system against published, peer-reviewed FTIR data from three structurally distinct nanomaterial systems, quantifying peak detection accuracy, positional error, and the scientific quality of the AI-generated spectral interpretations.

5. To develop a realistic business model and operational plan for SPECTRA AI, analysing its target market segments, value proposition, revenue strategy, competitive positioning, and pathway to commercial deployment as a sustainable scientific software product.

Structure of the Thesis

This thesis is organised into three chapters, each addressing a distinct dimension of the work.

Chapter I establishes the theoretical foundations of FTIR spectroscopy, covering the physical principles of infrared absorption, the instrumentation of Fourier transform spectrometers, the conventions of mid-infrared spectral interpretation, and the principal material classes encountered in characterisation studies.

Chapter II documents the engineering design and technical implementation of the SPECTRA AI intelligent assistant, detailing the software architecture, the peak detection algorithm and its parameterisation, the graphical user interface design, the AI report generation module and its prompt engineering strategy, and the multi-format export subsystem.

Chapter III presents the practical evaluation of the implemented tool, describing the interface in operational context and conducting a systematic experimental validation of the system's performance against three published nanomaterial spectrabiosynthesised Ag/AgCl nanoparticles, CuO@Tocopherol hybrid nanoparticles, and ZnO/$BaMg_2$ mixed-oxide nanoparticleswith a quantitative aggregate accuracy assessment and honest discussion of identified limitations.

CHAPTER I: Fundamentals of FTIR Spectroscopy

I. CHAPTER I: Fundamentals of FTIR Spectroscopy

I. 1. Introduction

Fourier Transform Infrared (FTIR) spectroscopy is a powerful analytical technique used to identify organic and inorganic materials by measuring their absorption of infrared light. It operates on the principle that molecules vibrate at specific frequencies, creating a unique molecular fingerprint for every substance. Unlike older dispersive methods, FTIR utilises a Michelson interferometer combined with the mathematical Fourier transformation to provide high-speed, high-resolution spectral data. This technique is indispensable for both qualitative and quantitative analysis, offering significant advantages in sensitivity and accuracy for modern chemical and materials research.

The present chapter establishes the theoretical and practical foundations on which the remainder of this dissertation is built. Section 1.1 situates FTIR within the broader landscape of material characterisation techniques. Section 1.2 develops the physical principles of infrared spectroscopy, the structure of a typical spectrum, the strategy by which functional groups are identified from band positions, and the practical sample-preparation methods used to record reliable spectra. Section 1.3 surveys the existing landscape of digital and automated analysis tools, identifies their limitations, and motivates the need for an intelligent, AI-augmented assistant the system whose design, implementation, and validation form the subject of subsequent chapters.

I. 2. Material Characterisation Techniques: An Overview

Modern materials research relies on a suite of complementary characterisation techniques, each probing a different physical property of the sample. No single technique provides a complete description of a material; instead, robust characterisation generally requires the combination of structural, morphological, thermal, optical, and chemical-bonding analyses. This section briefly situates Fourier Transform Infrared (FTIR) spectroscopy within this broader landscape and clarifies the specific information that it provides relative to other widely used methods [1].

X-ray diffraction (XRD) is the standard technique for probing crystalline structure. By measuring the angular distribution of X-rays scattered by the periodic atomic lattice, XRD enables phase identification, the determination of lattice parameters, and an estimation of

average crystallite size through line-broadening analysis. Its principal limitation is that it is largely insensitive to amorphous or short-range ordered phases, and it provides no direct information about chemical bonding or functional groups [2].

Scanning electron microscopy (SEM), frequently coupled with **energy-dispersive X-ray spectroscopy (EDX)**, provides high-resolution images of the sample surface together with localised elemental composition. SEM/EDX is therefore the technique of choice for assessing morphology, particle size, and elemental homogeneity. However, the sample must withstand the vacuum environment of the microscope, and conductive coatings are often necessary, neither of which yields information about the chemical bonds or functional groups present [3].

Thermogravimetric analysis (TGA) records the mass of a sample as a function of temperature under a controlled atmosphere. The resulting thermograms allow the quantitative determination of moisture and volatile content, the identification of decomposition steps, and an evaluation of thermal stability. TGA is destructive and does not provide chemical-bond-specific information, but it is invaluable for confirming compositional aspects suggested by spectroscopic methods [4].

Ultraviolet–visible (UV–Vis) spectroscopy measures electronic transitions induced by photons in the ultraviolet to visible range. It is widely used for determining optical band-gap energies of semiconductors, monitoring concentrations through the Beer–Lambert relation, and observing colour changes associated with chemical transformations. Its scope is restricted to materials whose electronic transitions fall within the accessible spectral window [5].

FTIR spectroscopy, in contrast to all of the above, probes the *vibrational* energy levels of chemical bonds. It is therefore uniquely well suited to the identification of functional groups and to the confirmation of molecular and supramolecular structure. FTIR is rapid, non-destructive in most sampling configurations, requires only minimal sample preparation, and is sensitive to both organic and inorganic chemistries. These attributes make FTIR a cornerstone of routine materials characterisation, complementary to rather than competitive with the techniques described above [6].

Table 1.1 summarises the principal complementary techniques discussed here, the property each one probes, the information it yields, and its main practical limitations. The remainder of this chapter focuses exclusively on FTIR, whose theoretical principles, spectral structure, and analytical capabilities provide the foundation for the intelligent assistant tool described in subsequent chapters.

Table I-1 Comparison of common material-characterisation techniques. FTIR is shown in bold to highlight its distinctive role in probing chemical bonding.

Technique	Property Probed	Information Obtained	Typical Limitations
XRD	Diffraction of X-rays by a crystalline lattice	Phase identification, crystallite size, lattice parameters	Insensitive to amorphous phases; bulk-averaged signal
SEM / EDX	Secondary-electron emission; X-ray fluorescence	Surface morphology, particle size, elemental composition	Vacuum required; conductive coating may be needed
TGA	Mass change as a function of temperature	Thermal stability, decomposition steps, water content	Destructive; provides no chemical-bond information
UV–Vis	Electronic transitions in the UV–visible range	Band-gap energy, optical absorption, concentration	Limited to materials with accessible electronic transitions
FTIR	Vibrational excitation by mid-infrared radiation	Functional groups, chemical bonds, molecular fingerprint	Requires IR-active modes; sample preparation can be delicate

I. 3. FTIR Spectroscopy

Fourier Transform Infrared (FTIR) spectroscopy represents the modern standard for performing infrared analysis, having largely replaced the older dispersive methods that dominated the field through the mid-twentieth century. The acronym “FTIR” refers to the mathematical operation the Fourier transform used to convert the raw interferometric signal into the readable spectrum that the analyst interprets. The technique is particularly prized for its ability to generate a *molecular fingerprint*: no two distinct molecular structures yield exactly the same infrared spectrum, which makes the method an exceptionally powerful tool for both identification and structural confirmation [7].

I. 4. Historical Context and the IR Spectral Range

The existence of infrared radiation was first recognised in 1800 by the British astronomer Sir Frederick William Herschel, who observed that a thermometer placed beyond the red end of a prism-dispersed solar spectrum recorded a higher temperature than thermometers placed within the visible range. This observation established the existence of an invisible form of radiant energy lying at wavelengths longer than red light, and laid the foundation for the entire field of infrared spectroscopy [8]. The systematic exploitation of this radiation for chemical

analysis followed only in the twentieth century, with the development of the dispersive infrared spectrometer in the 1940s and the Fourier-transform infrared spectrometer in the 1960s.

Modern infrared spectroscopy operates within a relatively narrow but information-rich portion of the electromagnetic spectrum. The infrared region as a whole extends from approximately 0.8 μm (immediately beyond the red end of the visible spectrum) to about 1000 μm, where it merges with the microwave region. By convention this range is partitioned into three sub-regions whose boundaries are defined by the molecular phenomena that they probe: the *near infrared* (NIR), the *mid infrared* (MIR), and the *far infrared* (FIR). Each sub-region carries different chemical information and is exploited by a different class of instrument and application [8].

The *near infrared* extends from roughly 0.8 to 2.5 μm, corresponding to wavenumbers between 13 000 and 4 000 cm^{-1}. Bands appearing in this region are predominantly overtones and combination bands of the fundamental vibrations that occur at lower energy; their intensities are typically one to two orders of magnitude weaker than the fundamentals. NIR spectroscopy is widely used for rapid, non-invasive quantitative analysis in industrial, agricultural, and pharmaceutical settings, where its ability to penetrate samples without preparation is particularly valued [9].

The *mid infrared* extending from 2.5 to 25 μm, equivalent to 4 000 to 400 cm^{-1} is by far the most widely used spectral window in analytical chemistry and is the range within which all the work described in this dissertation is conducted. Mid-IR radiation excites the fundamental vibrational modes of essentially every functional group of chemical interest, providing direct, unambiguous evidence of molecular structure. The mid-IR region is itself sometimes subdivided into a functional-group region (4000–1500 cm^{-1}) and a fingerprint region (1500–400 cm^{-1}) [10].

The *far infrared* spans approximately 25 μm to 1 mm, or 400 to 10 cm^{-1}. This region contains lattice vibrations of crystalline solids and the heavy-atom stretching modes of organometallic and inorganic complexes (notably metal–halide and metal–ligand vibrations). Far-IR work historically required specialised optical components transparent to long-wavelength radiation, although the development of broad-range diamond-window FTIR instruments has made the region considerably more accessible in recent years. Table 1.2 summarises the three sub-regions, their boundaries, and their principal analytical uses, and Figure 1.1 places them in the context of the broader electromagnetic spectrum [11].

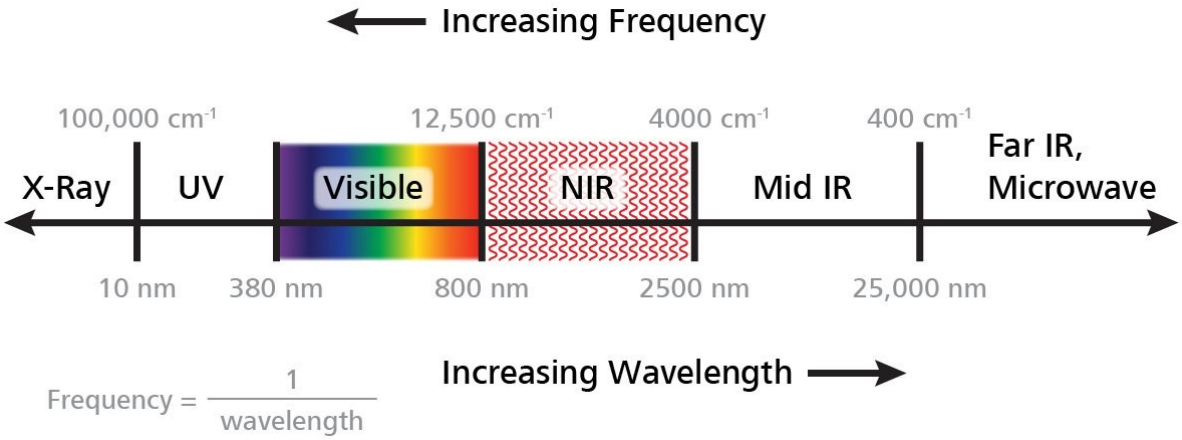


Figure I-1 Position of the infrared sub-regions within the electromagnetic spectrum. The mid-infrared range (4000–400 cm⁻¹) used in this dissertation is shown in bold [11].

Table I-2 Sub-regions of the infrared spectrum, their wavelength and wavenumber boundaries, and their principal analytical uses.

Sub-region	**Wavelength (µm)**	**Wavenumber (cm⁻¹)**	**Principal Use**
Near IR (NIR)	0.8 – 2.5	13 000 – 4 000	Overtones, combination bands; quantitative analysis
Mid IR (MIR)	2.5 – 25	4 000 – 400	Fundamental vibrations; standard FTIR range
Far IR (FIR)	25 – 1 000	400 – 10	Lattice vibrations; metal–ligand stretches

I. 5. Principle of Infrared Spectroscopy

The fundamental principle of infrared (IR) spectroscopy lies in the interaction between electromagnetic radiation and the chemical bonds within a sample. When infrared light is directed at a material, specific frequencies are absorbed by the molecules, causing transitions between quantised vibrational energy levels. This selective absorption gives rise to the characteristic molecular fingerprint that allows precise chemical identification of the analyte [12].

In modern analytical chemistry, the FTIR variant has become the standard implementation of the technique. Whereas older dispersive instruments separated incident light into its constituent wavelengths using a grating or prism and measured each wavelength sequentially,

FTIR measures all infrared frequencies simultaneously through a process known as *spectral encoding*, achieved by an optical device known as the Michelson interferometer [13].

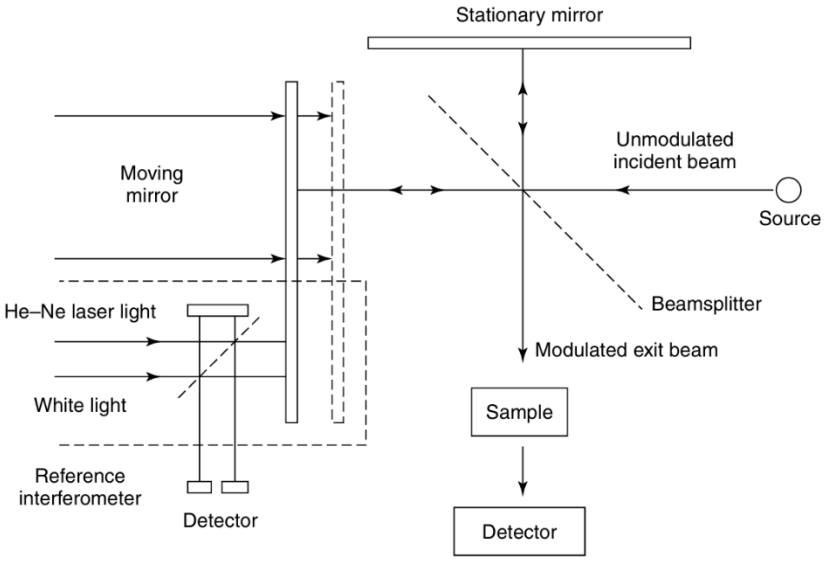


Figure I-2 Schematic diagram of a Michelson interferometer in an FTIR spectrometer [57].

I. 5. 1. The Optical Heart: The Michelson Interferometer

The core of the FTIR spectrometer is the interferometer, which is responsible for creating the encoded signal that the instrument subsequently decodes. As illustrated in Figure 1.2, the optical mechanism may be described in terms of four sequential stages. An infrared source first emits a beam containing a broad continuum of frequencies. This beam strikes a beamsplitter a semi-transparent device that reflects approximately half of the radiation toward a fixed mirror and transmits the remaining half toward a moving mirror. Because the moving mirror translates back and forth at constant velocity, the optical path travelled by the radiation reflected from it is continuously varying, while that travelled by the radiation reflected from the fixed mirror remains constant. When the two beams recombine at the beamsplitter, they interfere either constructively or destructively, depending on the instantaneous position of the moving mirror, and the resulting modulated beam is directed through the sample compartment toward the detector [13].

I. 5. 2. From Interferogram to Spectrum

The signal recorded by the detector is termed an *interferogram*: a complex, time-domain waveform that simultaneously contains information about every infrared frequency emitted by the source. As the modulated beam passes through the sample, the molecular absorption modifies the contribution of each frequency to the interferogram. The raw interferogram, however, is not directly interpretable it is the *sum* of all the spectral information rather than its dispersion across a wavenumber axis. The conversion from interferogram (time-domain) to spectrum (frequency-domain) is performed mathematically through the Fourier transform, and is executed in milliseconds by the instrument's acquisition computer [14]. The Fourier transform is therefore not merely a post-processing step but the central operation that gives the technique its name and its characteristic combination of speed and spectral resolution.

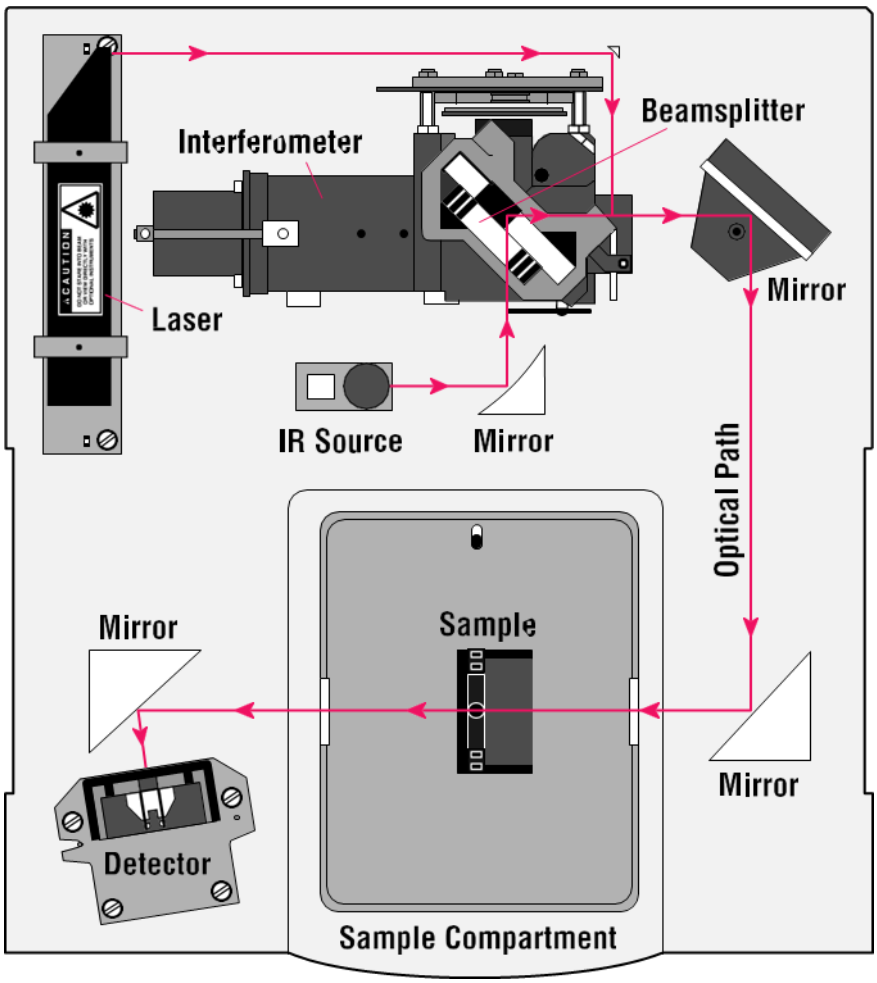


Figure I-3 A Schematic diagram showing different components and the optic path of a FT-IR spectrometer [58]

I. 5. 3. Key Technical Advantages

Three classical advantages of the Fourier-transform approach over the older dispersive method are routinely cited in the literature and account for the near-universal adoption of FTIR.

The *multiplex* (or Fellgett) advantage arises from the simultaneous measurement of all frequencies: a complete spectrum can be acquired in the time required by a dispersive instrument to scan a single resolution element, dramatically improving the signal-to-noise ratio for a given measurement time. The *throughput* (or Jacquinot) advantage stems from the absence of slits in the FTIR optical path, which permits a much larger fraction of the source radiation to reach the detector. Finally, the *Connes* advantage refers to the use of a stable internal reference laser to track the position of the moving mirror with sub-micron precision, ensuring exceptional wavenumber accuracy and reproducibility from one scan to the next [14].

I. 6. Molecular Vibrations

In the conceptual framework of vibrational spectroscopy, molecules are typically described using a *ball-and-spring* model: atoms are represented as point masses, and chemical bonds as Hookean springs of variable stiffness. Within this model, the atoms execute continuous, rhythmic oscillations whose characteristic frequencies depend on the stiffness of the spring (the bond force constant) and on the masses of the atoms involved. The mid-infrared region of the electromagnetic spectrum (roughly 4000 to 400 cm^{-1}) corresponds precisely to the energy range of these molecular vibrations, which is why infrared radiation is so effective at exciting them [15].

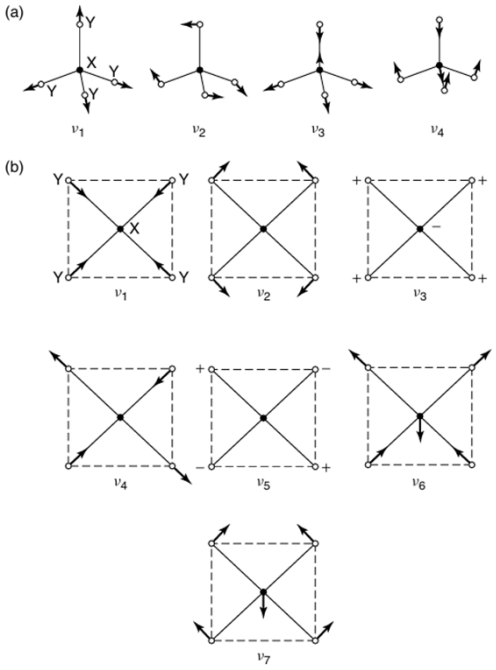


Figure I-4 Fundamental vibrational modes of polyatomic molecules, illustrating the principal stretching and bending mechanisms involved in infrared spectroscopy [16].

I. 6. 1. The Core Requirement: Dipole Moment Change

Not every molecular vibration can be detected by an FTIR spectrometer. For a vibration to be *infrared-active*, it must produce a change in the net dipole moment of the molecule during the vibrational motion. When the frequency of the incoming infrared radiation matches the natural frequency of such a vibration, the photon is absorbed and the molecule is promoted to a higher vibrational energy level. Vibrations that proceed without any change in dipole moment, such as the symmetric stretch of a homonuclear diatomic molecule, are intrinsically infrared-inactive and produce no absorption band, regardless of the intensity of the incident radiation [15].

I. 6. 2. Fundamental Modes of Vibration

Molecular vibrations are conventionally classified into two principal categories: *stretching* and *bending* (or deformation) modes. Stretching vibrations involve a continuous variation in the bond length along the internuclear axis and may be either symmetric in which both atoms move toward and away from the centre of the bond simultaneously or asymmetric, in which one atom moves toward the centre while the other moves away. Stretching modes generally require greater energy than bending modes and consequently appear at higher wavenumbers in the spectrum [17].

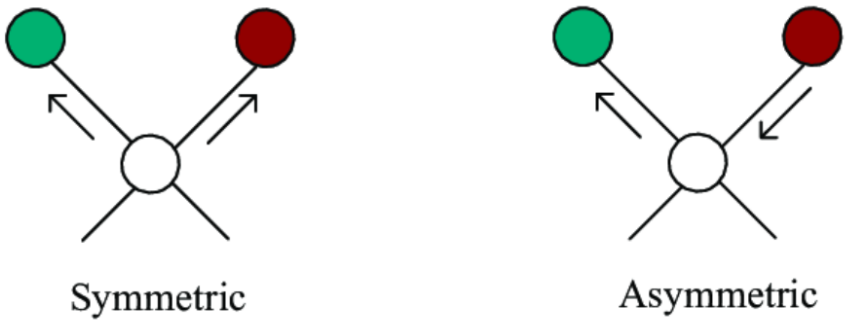


Figure I-5 Stretching vibrations [18]

Bending vibrations involve a change in the angle between two bonds rather than in the bond length itself, and they typically occur at lower wavenumbers. Four sub-types are usually distinguished: *scissoring*, in which two atoms move toward and away from each other within a common plane; *rocking*, in which an atomic group swings rigidly back and forth in the molecular plane; *wagging*, in which the atoms move synchronously up and down out of the molecular plane; and *twisting*, in which the atoms rotate about the bond connecting them to the remainder of the molecule. Figure I.5 illustrates these motions schematically and serves as the conceptual reference for the assignment of bands in the chapters that follow [19].

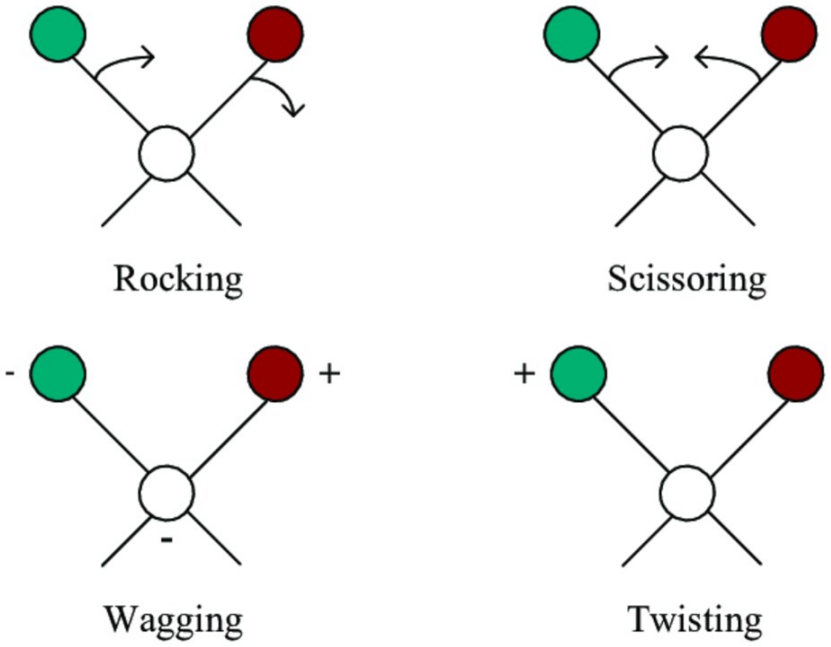


Figure I-6 Bending vibrations [18]

I. 6. 3. Mathematical Model: Hooke's Law

To predict the spectral position of a vibration, the bond is modelled as a classical *harmonic oscillator*. According to Hooke's law, the wavenumber of vibration ν is related to the bond force constant k and the reduced mass μ of the two oscillating atoms through the relation

$$\bar{\nu} = \frac{1}{2\pi c}\sqrt{\frac{k}{\mu}} \qquad ()$$

$$\text{where} \quad \mu = \frac{m_1 \times m_2}{m_1 + m_2} \qquad ()$$

In this expression ν is the wavenumber expressed in cm^{-1}, c is the speed of light, k is the force constant of the bond (a measure of bond stiffness, with units of N m^{-1}), and μ is the reduced mass of the two-atom system, with m_1 and m_2 the masses of the individual atoms. The relation embodies a simple but powerful rule of vibrational spectroscopy: stiffer bonds (larger k) and lighter atoms (smaller μ) vibrate at higher wavenumbers. This is why O–H, N–H, and C–H stretches all appear above 2800 cm^{-1}, while metal–oxygen vibrations of heavy lattice atoms typically lie below 800 cm^{-1} [20].

I. 6. 4. Degrees of Freedom and Normal Modes

Every polyatomic molecule possesses a specific number of independent vibrational degrees of freedom, referred to as *normal modes*. For a non-linear molecule containing N

atoms, the number of normal modes is $3N - 6$, while for a linear molecule the count reduces to $3N - 5$ because the rotation about the molecular axis carries no inertia. Table 1.3 summarises this counting rule with representative examples. In linear CO_2 ($N = 3$), four fundamental vibrations are predicted; however, the symmetric stretch produces no change in dipole moment and is therefore IR-inactive, while the asymmetric stretch and the doubly degenerate bending mode are both active and contribute observable bands to the spectrum [21].

Table I-3 Determining the molecular degrees of vibrational freedom from the number of atoms N and molecular geometry.

Molecular Geometry	Calculation Formula	Example
Linear	$3N - 5$	Carbon dioxide (CO_2)
Non-linear	$3N - 6$	Water (H_2O)

I. 6. 5. Complicating Factors in Real Spectra

Although fundamental vibrations dominate the appearance of an infrared spectrum, real spectra commonly contain additional features that enrich and sometimes obscure the molecular fingerprint. *Overtones* appear at approximately integer multiples of the fundamental frequency (for instance, near $2\nu_1$ for a first overtone) and arise from transitions to higher vibrational levels with reduced probability. *Combination bands* occur when a single photon excites two distinct vibrations simultaneously, and consequently appear at the sum of the two fundamental frequencies. *Fermi resonance* is observed when a fundamental vibration and an overtone (or combination band) lie close in energy and possess matching symmetry: the two states mix quantum-mechanically, shifting their apparent frequencies and redistributing intensity between them [22]. Awareness of these phenomena is essential for accurate spectral interpretation, as they can otherwise be mistakenly assigned to additional functional groups.

I. 7. FTIR Spectrum Structure

An infrared spectrum is a graphical representation of how a molecule interacts with infrared light, plotted on a two-dimensional coordinate system that translates physical absorption events into visual data. Understanding this graphical framework is a prerequisite for the chemical interpretation of spectra, since each axis encodes a different type of physical information [23].

I. 7. 1. The Graphical Framework

The horizontal axis of an infrared spectrum reports the *wavenumber* ν, expressed in reciprocal centimetres (cm^{-1}). Wavenumber is directly proportional to photon energy, so high-

energy vibrations appear on the left of the spectrum (typically beginning at 4000 cm^{-1}) while lower-energy vibrations appear toward the right (extending to approximately 400 cm^{-1}). The choice of wavenumber over wavelength is conventional in vibrational spectroscopy because of the resulting linearity in energy and its convenience for the harmonic-oscillator framework discussed above [24].

The vertical axis reports the *intensity* of the absorption and may be expressed in either of two related ways. *Transmittance* (*T*) gives the fraction of incident light that has passed through the sample, with absorption bands appearing as downward deviations from a baseline near 100 %. *Absorbance* ($A = -\log_{10} T$) gives the logarithm of the inverse transmittance, with absorption bands appearing as upward peaks. The absorbance representation is preferred for quantitative analysis because it follows the linear Beer–Lambert relation between concentration and signal, while transmittance is more commonly encountered in qualitative materials characterisation [25].

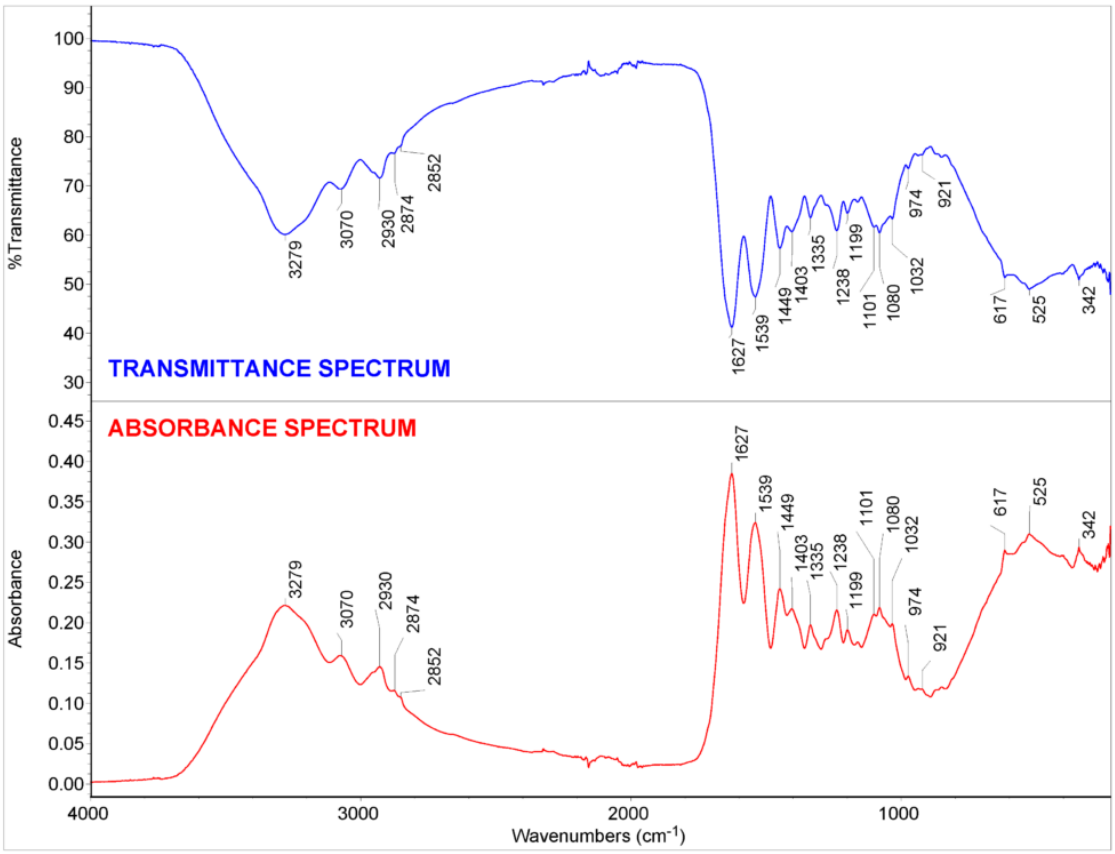


Figure I-7 Comparison between the transmittance and absorbance representations of an FTIR spectrum [26].

I. 7. 2. Spectral Regions: The Diagnostic and the Fingerprint

A standard mid-infrared FTIR spectrum is conventionally divided into two principal regions that serve different diagnostic purposes, summarised in Table 1.4. The *functional-group region*, spanning approximately 4000–1500 cm^{-1}, contains relatively few but highly diagnostic peaks, each characteristic of a specific functional group such as O–H, C=O, or N–H. Bands in this region are largely transferable across molecules, so a spectroscopist can recognise the building blocks of a molecule from a few key absorptions. The *fingerprint region*, between 1500 and 400 cm^{-1}, is by contrast densely populated with bending and skeletal vibrations whose individual assignments are difficult, but whose overall pattern is unique to a given compound. The fingerprint region is therefore most useful for confirmatory identification by comparison with reference spectra rather than for de-novo functional-group assignment [27].

Table I-4 Major regions of the mid-infrared spectrum and their primary diagnostic significance.

Region Name	Wavenumber Range (cm^{-1})	Primary Significance
Functional Group Region	4000 – 1500	Identification of specific chemical groups (e.g., O–H, C=O, N–H)
Fingerprint Region	1500 – 400	Unique patterns from complex bending and skeletal vibrations

I. 7. 3. Anatomy of a Spectral Peak

Three physical attributes of an absorption band carry chemical information. The *position* of the band its wavenumber reflects the energy of the corresponding vibration and is determined by bond strength and atomic mass through Hooke's law. The *intensity* of the band reflects the magnitude of the dipole-moment change during the vibration and the concentration of the absorbing species in the sample; intensely polar bonds such as C=O typically produce strong absorption bands, while symmetric, weakly polar vibrations may be very weak or even absent. The *shape* of the band sharp, broad, or asymmetric encodes information about the molecular environment: hydrogen-bonded O–H groups, for example, produce characteristically broad and rounded bands whose width is itself diagnostic of the hydrogen-bond strength [28].

I. 8. Functional-Group Identification

The principal practical power of FTIR spectroscopy lies in its ability to act as a diagnostic tool for identifying the functional groups within a molecule. Each chemical bond C=O, O–H, C–H, and so on behaves as a distinct harmonic oscillator that absorbs infrared radiation at a characteristic frequency that varies only modestly with the chemical environment. By systematically analysing the wavenumbers, intensities, and shapes of the absorption bands, the analyst can reconstruct the molecular architecture of an unknown sample, often with considerable confidence [29].

I. 8. 1. Diagnostic Strategy: Two Complementary Regions

For practical identification, the spectrum is treated as two complementary zones. The functional-group region (4000–1500 cm^{-1}) is the natural starting point because its bands are clean, comparatively isolated, and characteristic of specific bond types regardless of molecular context a band near 1715 cm^{-1}, for instance, almost invariably indicates a carbonyl (C=O) group. Once the major functional groups have been tentatively assigned from this region, the

fingerprint region (1500–400 cm^{-1}) is consulted for confirmatory identification: rather than attempting peak-by-peak interpretation, the entire fingerprint pattern is compared against reference spectra of candidate compounds, much as a forensic fingerprint is matched [30].

I. 8. 2. Key Functional Groups and Their Frequencies

Identification depends on familiarity with the characteristic absorption ranges of common bonds. Table 1.5 summarises the most frequently encountered functional groups in materials analysis, organised by chemical family alkanes, alkenes and aromatics, alkynes, alcohols and phenols, carbonyl compounds, amines, halogenated aliphatics, and nitriles. For each group the table gives the wavenumber range, the principal vibrational mode (using IUPAC notation: ν for stretching, δ for in-plane bending, γ for out-of-plane bending; subscripts s and as denote symmetric and asymmetric variants), and a short note on the typical band character. The ranges are deliberately presented as intervals rather than single values because the precise band position within each range depends on the local chemical environment a fact that the AI-assisted interpretation system described in Chapter II must explicitly accommodate [1].

Table I-5 Characteristic infrared absorption frequencies of common functional groups encountered in materials analysis, organised by chemical family.

Bond / Group	**Wavenumber (cm^{-1})**	**Vibration**	**Notes**
Alkanes			
–CH_3 (methyl)	2970 – 2950	ν_{as}(C–H)	Asymmetric stretch
–CH_3	2880 – 2860	ν_{s}(C–H)	Symmetric stretch
–CH_3	1470 – 1430	δ_{s}(C–H)	Symmetric bend
–CH_3	1380 – 1370	δ_{as}(C–H)	Asymmetric bend
>CH_2 (methylene)	2935 – 2915	ν_{as}(C–H)	Asymmetric stretch
>CH_2	2865 – 2845	ν_{s}(C–H)	Symmetric stretch
>CH_2	1485 – 1445	δ(C–H)	Scissoring
>CH– (methine)	2900 – 2880	ν(C–H)	Stretching
O–CH_3 (methoxy)	2850 – 2815	ν(C–H)	Methyl ether
Alkenes and Aromatics			
=C–H (vinyl)	3095 – 3075	ν(C–H)	Sharp, medium
C=C (aliphatic)	1680 – 1640	ν(C=C)	Variable intensity
C=C (aromatic)	1615 – 1450	ν(C=C)	Multiple bands; benzene ring
Ar–H	3130 – 3070	ν(C–H)	Aromatic C–H stretching
Ar–H	900 – 670	γ(C–H)	Out-of-plane bend; substitution pattern
Alkynes			
C≡C	2300 – 2100	ν(C≡C)	Sharp, weak to medium
≡C–H	3320 – 3310	ν(C–H)	Sharp, characteristic
Alcohols and Phenols			
O–H (free)	3650 – 3600	ν(O–H)	Sharp, narrow

O–H (H-bonded)	3400 – 3200	ν(O–H)	Strong and broad
C–O	1260 – 1000	ν(C–O)	Strong; primary, secondary, tertiary differ
Carbonyl Compounds			
C=O (aldehyde)	1740 – 1725	ν(C=O)	Strong, sharp
C=O (ketone)	1725 – 1705	ν(C=O)	Strong, sharp
C=O (carboxylic ac.)	1725 – 1700	ν(C=O)	Often broadened by H-bonding
C=O (ester)	1750 – 1725	ν(C=O)	Strong, sharp
C=O (amide)	1680 – 1630	ν(C=O)	Amide I band
Amines			
N–H (primary)	3400 – 3380	ν(N–H)	Two bands; medium
N–H (secondary)	3360 – 3310	ν(N–H)	Single band; medium
N–H (primary)	1650 – 1590	δ(N–H)	Bending
C–N	1090 – 1020	ν(C–N)	Primary amines
Aliphatic Halogenated Compounds			
C–F	1150 – 1000	ν(C–F)	Strong
C–Cl	800 – 700	ν(C–Cl)	Strong
C–Br	700 – 600	ν(C–Br)	Strong
C–I	600 – 500	ν(C–I)	Strong
Nitriles			
C≡N	2260 – 2210	ν(C≡N)	Sharp; medium intensity

I. 8. 3. Factors Influencing Peak Position

Spectral interpretation is rarely a mechanical exercise of matching wavenumbers to a table, because the precise position of a functional-group band can shift appreciably with the molecular environment. Two effects are particularly important. *Hydrogen bonding*, prevalent in O–H and N–H groups, weakens the covalent bond by partial proton sharing with a neighbouring electronegative atom; the result is a shift of the band toward lower wavenumber accompanied by significant broadening of the band. *Conjugation* the alternation of single and double bonds, or the proximity of a multiple bond to an aromatic ring delocalises the π-electrons across multiple atoms, lowering the effective force constant of the C=O or C=C bond and consequently shifting its absorption band toward lower wavenumber. Familiarity with these effects is essential for distinguishing genuine functional-group differences from environmental shifts [31].

I. 8. 4. The Role of Peak Intensity

Band intensity carries information complementary to band position. The integrated absorption intensity is proportional to the square of the change in dipole moment during the vibration, so highly polar bonds such as C=O produce some of the most intense bands in any infrared spectrum, while symmetric bonds in highly symmetric environments for instance, the

central C=C bond of an symmetrically substituted ethylene may be extremely weak or even fully IR-inactive. Combined with peak position and shape, intensity allows the spectroscopist to move from a raw spectrum to a chemically meaningful functional-group profile, forming the basis for advanced material characterisation [31].

I. 9. Applications in Material Characterisation

Infrared spectroscopy has established itself as a versatile and indispensable tool across materials science, industry, and biological research. By providing a unique molecular fingerprint, FTIR enables the identification and structural characterisation of an extraordinarily diverse range of materials, from simple organic compounds to complex biological tissues and engineered industrial polymers [32].

In the *industrial and environmental* context, FTIR is critical for ensuring product quality and consistency in pharmaceutical manufacturing, where it is routinely used for the characterisation of active ingredients and the detection of crystalline polymorphs whose presence affects dosage stability. In environmental monitoring, the technique can detect gas-phase pollutants at concentrations as low as 10 ppb, allowing real-time monitoring of trace contaminants in air and water [33].

In *polymer and coatings* science, FTIR is the principal technique for the identification of unknown resins and synthetic fibres and for the study of degradation processes. Specialised sampling configurations notably attenuated total reflectance (ATR) and reflectance modes permit the analysis of surface coatings and treatments without damaging the underlying substrate, which is particularly valuable when only a small sample is available or when destructive analysis is precluded by the cost or rarity of the speciment [34].

Modern advances have extended the reach of FTIR into *biological and medical* fields. Subtle changes in the infrared spectra of tissues can serve as indicators of disease, while the quantitative analysis of characteristic protein and nucleic-acid bands provides insight into the secondary structure of proteins and the composition of nucleic acids in cell extracts and biological matrices. Such applications are inherently more demanding than routine industrial use because the signals are weaker, the matrices more complex, and the chemical interpretation correspondingly more nuanced [35].

Finally, the integration of infrared spectroscopy with optical microscopy has produced *infrared microscopy and imaging*, a class of techniques that allow the spectroscopic

characterisation of individual particles, microscopic contaminants, or sub-millimetre regions of a heterogeneous sample. The resulting chemical maps can be used to identify forensic micro-traces or to localise compositional anomalies across the surface of an industrial component. Table 1.6 summarises the principal application domains of FTIR in materials characterisation and the practical benefits each offers.

Table I-6 Common applications of FTIR in material characterisation and the principal benefit each offers.

Field	Typical Application	Key Benefit
Polymers	Identifying unknown resins; monitoring degradation	Quick and non-destructive
Forensics	Analysing paint chips, fibres, and microscopic traces	High sensitivity for micro-samples
Environment	Detecting gas pollutants (e.g., CO_2, vapours)	Real-time monitoring
Biology	Identifying microbial cells; tissue diagnostics	Distinctive spectral fingerprints
Pharmaceuticals	Polymorph detection and drug-quality control	Essential for dosage stability

In summary, the capacity of FTIR to provide rich molecular information rapidly and non-destructively makes it a cornerstone of modern materials characterisation. The reliability of the resulting spectra, however, depends critically on the manner in which the sample is presented to the instrument; for that reason the practical methods of sample preparation are reviewed in the following subsection before turning to the limitations of existing analysis software.

I. 10. Sample-Preparation Techniques

The quality of an FTIR spectrum is determined as much by sample preparation as by the spectrometer itself. Inappropriate sample handling can produce broad, distorted, or saturated bands that defeat any subsequent automated analysis, irrespective of the sophistication of the algorithms applied to the recorded spectrum. Five sample-preparation strategies dominate routine practice, each appropriate to a different physical state and analytical objective; the choice between them is determined by sample form, spectral region of interest, and the trade-off between preparation effort and reproducibility [31].

I. 10. 1. Liquid Films

For neat liquid samples, a small drop is sandwiched between two polished sodium chloride (NaCl) discs to form a thin film of approximately 10–25 μm thickness. NaCl is transparent throughout the 4000–625 cm^{-1} range, which encompasses the entire functional-group region

and most of the fingerprint region. The plates are extremely fragile and hygroscopic and must therefore be stored in a desiccator, handled only by their edges, and never exposed to aqueous samples. After analysis, residual sample is removed with anhydrous ethanol or another non-aqueous solvent. Cloudy or pitted plates produce broadened bands and reduced overall transmission and should be repolished or replaced before use [31].

I. 10. 2. Solutions

When a solid analyte is soluble in a suitable solvent, the resulting solution may be analysed in a sealed solution cell consisting of two NaCl windows separated by a metal spacer of defined thickness, typically 0.05–0.5 mm. The principal challenge is solvent selection: every common solvent has its own infrared absorption spectrum, and bands of the analyte that overlap with strong solvent bands cannot be reliably observed. Carbon tetrachloride, chloroform, and carbon disulfide are widely used because each is transparent across complementary regions of the mid-infrared range; no single solvent is transparent throughout. A reference spectrum of the pure solvent must be recorded under identical conditions and subtracted from the sample spectrum to remove the solvent contribution. Where aqueous solubility is unavoidable, calcium fluoride (CaF_2) cells must be substituted for NaCl because of the latter's water sensitivity [31].

I. 10. 3. Nujol Mulls

Solid samples that cannot be dissolved without chemical alteration may be examined as a *mull* a fine suspension of the powdered solid in a chemically inert mineral oil. Nujol, a high-boiling paraffin, is the standard mulling agent. A small quantity of finely ground sample is combined with a drop of Nujol in an agate mortar, the resulting paste is spread between two NaCl windows, and the spectrum is recorded. Nujol itself absorbs strongly in the C–H stretching region (around 2900 cm^{-1}) and the C–H bending region (around 1450 and 1375 cm^{-1}), so any analyte bands in these intervals are obscured. To address this limitation, a complementary mull in a fluorinated grease (Fluorolube) is sometimes prepared and the two spectra combined to provide complete spectral coverage [36].

I. 10. 4. Pressed KBr Discs

The most widely used preparation method for solid samples in academic and industrial laboratories is the pressed potassium bromide (KBr) disc. A small mass of finely ground sample (typically 1–3 mg) is mixed thoroughly with 100–300 mg of dry, IR-grade KBr, and the mixture is compressed in a hydraulic press at approximately 10 tonnes for several minutes. Under this pressure the KBr undergoes cold-flow plastic deformation and forms a transparent, glass-like

disc within which the analyte particles are uniformly dispersed. KBr is transparent across the entire 4000–400 cm^{-1} range, so the resulting spectrum suffers no solvent or matrix interference of the kind that affects mulls. A residual band near 3450 cm^{-1}, attributable to adsorbed water, is commonly observed and may be minimised by drying the KBr at 110 °C and grinding under low-humidity conditions [37]. Solid-state spectra obtained in KBr discs frequently exhibit a greater number of resolved bands than the corresponding solution spectra, because intermolecular interactions in the crystalline state most notably hydrogen bonding contribute additional features that aid compound identification.

I. 10. 5. Attenuated Total Reflectance (ATR)

The most significant practical advance in FTIR sampling over the past three decades is the attenuated total reflectance (ATR) accessory. Unlike the transmission methods described above, ATR exploits a fundamentally different optical principle: the infrared beam is directed into a high-refractive-index crystal at an angle that produces total internal reflection at the crystal–sample interface. At each reflection point, an *evanescent wave* extends a short distance typically 0.5 to 2 μm into the sample placed in optical contact with the crystal. The molecular absorption of the analyte attenuates this evanescent wave; when the beam ultimately exits the crystal, the reduction in intensity at each wavelength constitutes the ATR spectrum [37].

ATR has revolutionised routine FTIR analysis because it eliminates almost all of the sample-preparation effort associated with transmission methods. Liquids are placed directly on the crystal; pastes and gels are spread thinly across its surface; powders and solid samples are pressed onto the crystal using an integrated pressure clamp. No grinding, no hydraulic press, no solvent selection is required, and the spectrum is recorded in seconds. The principal limitation is that the path length of interaction is fixed by the evanescent-wave penetration depth and is therefore wavelength-dependent: longer-wavelength radiation penetrates more deeply, producing apparent intensity differences relative to transmission spectra that must be corrected algorithmically before quantitative comparisons are made.

The choice of ATR crystal is determined by spectral range, chemical compatibility, and durability. Zinc selenide (ZnSe) is the most common general-purpose crystal but is restricted to a working pH range of 5–9 and is sensitive to abrasive samples. Germanium (Ge) is chemically robust across pH 1–14 and is the preferred choice for strong acids and bases, but its high refractive index limits penetration depth. Diamond is the most durable and chemically inert option, accommodates the widest spectral range, and is the standard choice for hard or

abrasive solids; its higher initial cost is offset by extended operating life. The principal optical and chemical characteristics of the most common crystal materials are summarised in Table 1.7.

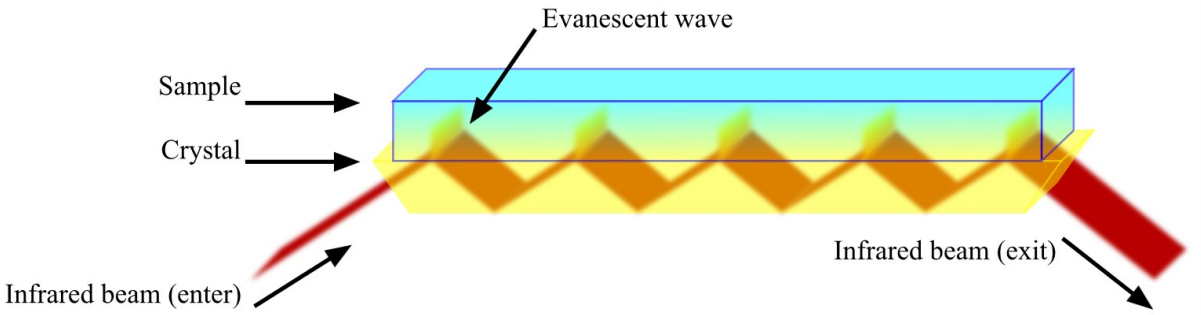


Figure I-8 Operating principle of an attenuated total reflectance (ATR) accessory: the infrared beam undergoes total internal reflection within the crystal, and the evanescent wave penetrates into the sample placed in optical contact with the upper surface [40].

Table I-7 Common crystal materials used in attenuated total reflectance (ATR) accessories, with spectral range and refractive index.

Crystal Material	**Wavenumber Range (cm^{-1})**	**Refractive Index**	**Notes**
Zinc selenide (ZnSe)	20 000 – 500	2.43	pH 5–9; common
Zinc sulfide (ZnS)	22 000 – 750	2.25	Lower cost
Germanium (Ge)	5 000 – 600	4.01	pH 1–14; robust
Silicon (Si)	10 000 – 100	3.42	Far-IR access
Diamond	45 000 – 10	2.40	Most durable; broad range

Across all five preparation methods, a quantitative criterion for spectral acceptability is conventionally applied: the strongest absorption band in the recorded spectrum should reach a transmittance of approximately 2–5 %, corresponding to an absorbance of $A \approx 1.3$ from the relation $A = 2 - \log_{10}(\%\ T)$. Bands that exceed this absorbance value lie beyond the linear response range of most detectors and produce truncated, unreliable peaks; the sample must in such cases be diluted, the path length reduced, or for ATR the contact area decreased before the spectrum is re-recorded [37]. The choice of preparation method therefore implicitly determines the analytical workflow that follows: ATR-recorded spectra of solids and liquids feed the AI-driven assistant developed in Chapter II without further reduction, whereas

transmission spectra recorded as KBr discs or solution films may require manual baseline correction or normalisation before they can be processed automatically.

I. 11. Digital and Automated Spectral Analysis

The evolution of spectroscopy from manual interpretation toward digital and increasingly automated processing has transformed the practice of materials characterisation. Modern instruments produce vast quantities of high-resolution data, and modern users both academic and industrial increasingly expect the software environment surrounding the instrument to deliver not only the raw spectrum but also a degree of structured interpretation. This section surveys the existing landscape of digital spectral-analysis tools, identifies their persistent limitations, and motivates the development of a new class of *intelligent* assistants based on contemporary artificial-intelligence techniques.

I. 12. Existing Digital Tools

The current ecosystem of FTIR analysis is dominated by integrated software environments designed to manage the complexity of modern digital data. At the lowest level sit *Fast Fourier Transform (FFT) algorithms* ***Figure I. 9***, the digital engines that convert the raw interferogram into the readable wavenumber-domain spectrum almost instantaneously and that are now sufficiently mature to be regarded as a solved problem. Built on top of the FFT layer are *commercial spectral libraries*: digital databases containing tens of thousands of reference spectra against which an unknown spectrum can be matched through automated search-and-match routines. The third layer comprises *quantitative software suites*, which apply the Beer–Lambert law and chemometric regression methods to determine the concentration of specific components in mixtures. Finally, *data-management systems* provide the infrastructure for the storage, statistical analysis, and archival of spectral data, ensuring traceability in regulated industrial settings [37].

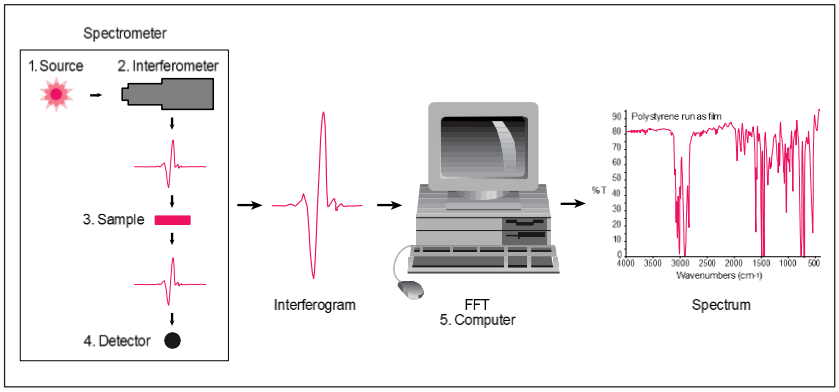


Figure I-9 The lowest level sit Fast Fourier Transform (FFT) algorithms [38]

I. 13. Limitations of Current Tools

Despite the maturity and commercial penetration of the digital tools surveyed above, several persistent technical and practical bottlenecks remain. The first is *sensitivity to matrix effects*: automated identification routines often struggle to detect trace components present at concentrations below 0.01 % when these components are embedded in a complex matrix that itself absorbs in the same spectral region. The second is *library dependency*: the accuracy of any search-and-match identification is bounded by the diversity and completeness of the reference library, with the consequence that compounds absent from the library cannot be identified at all, regardless of how well-resolved their spectra may be. The third limitation is the inability of standard tools to *correct for physical interferences* such as variations in sample orientation, particle size, or scattering, all of which can distort the recorded spectrum in ways that defeat purely pattern-based analysis [37].

Table I-8 Comparison between manual interpretation and current digital tools for FTIR spectral analysis.

Feature	Manual Interpretation	Current Digital Tools
Speed	Slow; depends on operator expertise	Rapid, real-time processing
Consistency	Subjective; prone to human error	Highly consistent and repeatable
Mixture analysis	Very difficult	Requires extensive calibration
Data storage	Physical records and printed plots	Digital databases and archives
Chemical insight	Strong contextual reasoning by expert	Pattern matching only; no reasoning

I. 14. The Need for an Intelligent Assistant

The gap between raw digital processing and meaningful chemical insight motivates the development of a new generation of *intelligent assistants* for spectroscopy. Unlike the standard software described above, an intelligent assistant would address several interconnected needs. First, it would provide *contextual reasoning*, moving beyond simple pattern matching to take into account the chemical environment and the plausibility of candidate assignments, thereby reducing the rate of false positives. Second, it would automate or substantially shorten the laborious calibration phases that currently precede the analysis of complex mixtures, replacing manual chemometric model construction with adaptive machine-learning approaches. Third, it would offer some degree of *error correction*, recognising and compensating for spectral distortions caused by sample-preparation artefacts or by minor instrumental drift. Fourth and arguably most consequentially it would serve as a *bridge for non-expert users*, generating guided interpretations that currently require years of specialised training and lowering the barrier to adoption in laboratories that lack a resident vibrational spectroscopist [39].

The recent advent of large language models (LLMs) has transformed the practical feasibility of such an assistant. Whereas earlier rule-based or expert-system approaches required the laborious encoding of spectroscopic heuristics into programs, modern LLMs can be prompted to produce structured, journal-quality scientific text directly from a list of detected peaks, provided that the prompt is carefully engineered to enforce scientific accuracy and the appropriate stylistic conventions. The combination of robust digital peak detection a task at which classical signal-processing algorithms excel with LLM-driven interpretive text generation defines the architectural strategy adopted in this work.

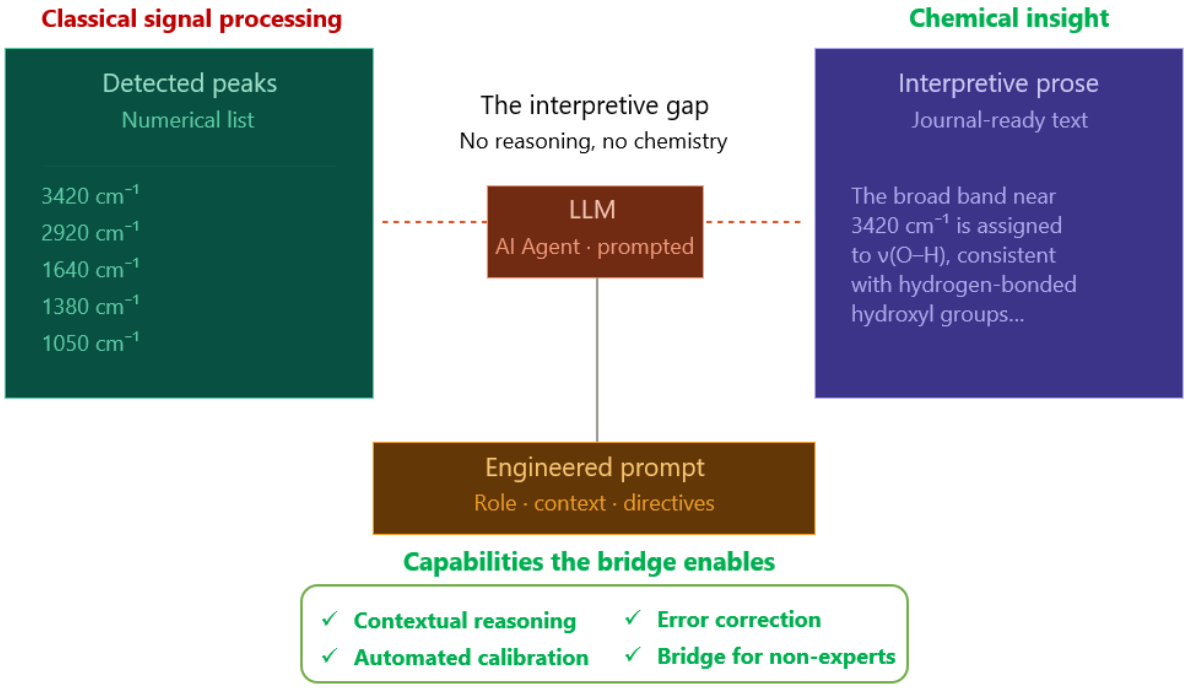


Figure I-10 Conceptual framework for an AI-driven FTIR spectral interpretation pipeline, integrating classical digital signal processing with large-language-model-based scientific text generation.

In conclusion, while digital tools have effectively digitised the molecular fingerprint of matter, the next frontier in vibrational spectroscopy lies in the integration of artificial intelligence to overcome the analytical and interpretive limitations of current software and to simplify the path from raw data to scientific insight. Chapter II details the design and development of such an intelligent assistant a software tool that combines a robust prominence-based peak-detection algorithm with a Google Gemini-powered AI report-generation module to produce publication-ready FTIR Results-and-Discussion text directly from a CSV spectrum file.

CHAPTER II: Design and Development of the Intelligent Assistant Tool

II. CHAPTER II: Design and Development of the Intelligent Assistant Tool

II. 1. Introduction and Design Objectives

The development of a software tool capable of autonomously interpreting Fourier Transform Infrared (FTIR) spectra and producing publication-ready scientific reports represents a significant intersection between analytical chemistry, signal processing, and artificial intelligence. This chapter provides a comprehensive account of the engineering decisions, software architecture, algorithms, and integration strategies employed in constructing such a system. The tool was designed to serve the needs of materials scientists and chemical engineers who require efficient, reproducible, and accurate spectral characterisation without reliance on expensive commercial software.

The motivation for building a custom tool, rather than extending an existing commercial package such as OPUS (Bruker), Spectrum (PerkinElmer), or OMNIC (Thermo Fisher), arose from three concrete limitations identified during the preparatory phase of this work. First, every commercial package examined enforces a closed-format binary file structure that prevents scripted batch processing without paid scripting add-ons. Second, none of the surveyed packages incorporates any form of generative AI for the production of interpretive text; spectral interpretation remains an entirely manual task performed by the operator after the data have been visualised. Third, the per-seat licence costs of these packages place them out of reach of many academic groups in developing economies, including the laboratory in which this work was conducted. A free, open, and AI-augmented alternative therefore addresses a real and present need in the materials characterisation community.

The principal design objectives established at the outset of this work were threefold. First, the system must be capable of reliable automated peak detection across a wide range of material types, including metal oxides, polymers, composites, and bio-derived compounds. Second, it must provide a graphical user interface (GUI) sufficiently intuitive for non-specialist users while preserving the analytical flexibility demanded by expert practitioners. Third, and most distinctively, the system must leverage a state-of-the-art large language model (LLM) to generate structured, journal-quality scientific text from the detected spectral data, dramatically reducing the time required to draft Results and Discussion sections for manuscript submission.

A fourth, cross-cutting requirement was that the entire system should be implementable using only free and open-source components and freely available cloud services. This requirement reflects a deliberate accessibility commitment: a tool that requires costly software

dependencies merely shifts the affordability problem rather than solving it. The selected technology stack Python with its scientific libraries, the Tkinter GUI framework bundled with the Python standard distribution, and the free tier of the Google Gemini API fully satisfies this constraint.

The following sections detail each component of the system in sequence: the overall architecture, the data ingestion pipeline, the peak detection algorithm and its parameterisation, the manual peak management subsystem, the visualisation engine, the graphical user interface, the AI report generation module and its prompt engineering methodology, the export and document generation subsystem, and a concluding summary of the technical specifications.

II. 2. Overall System Architecture

The intelligent assistant was designed according to a modular, layered architecture in which each functional component operates independently while communicating through well-defined interfaces. This approach facilitates maintenance, future extension, and the substitution of individual modules without requiring modification of the broader system. Figure II.1 provides a schematic overview of the system architecture.

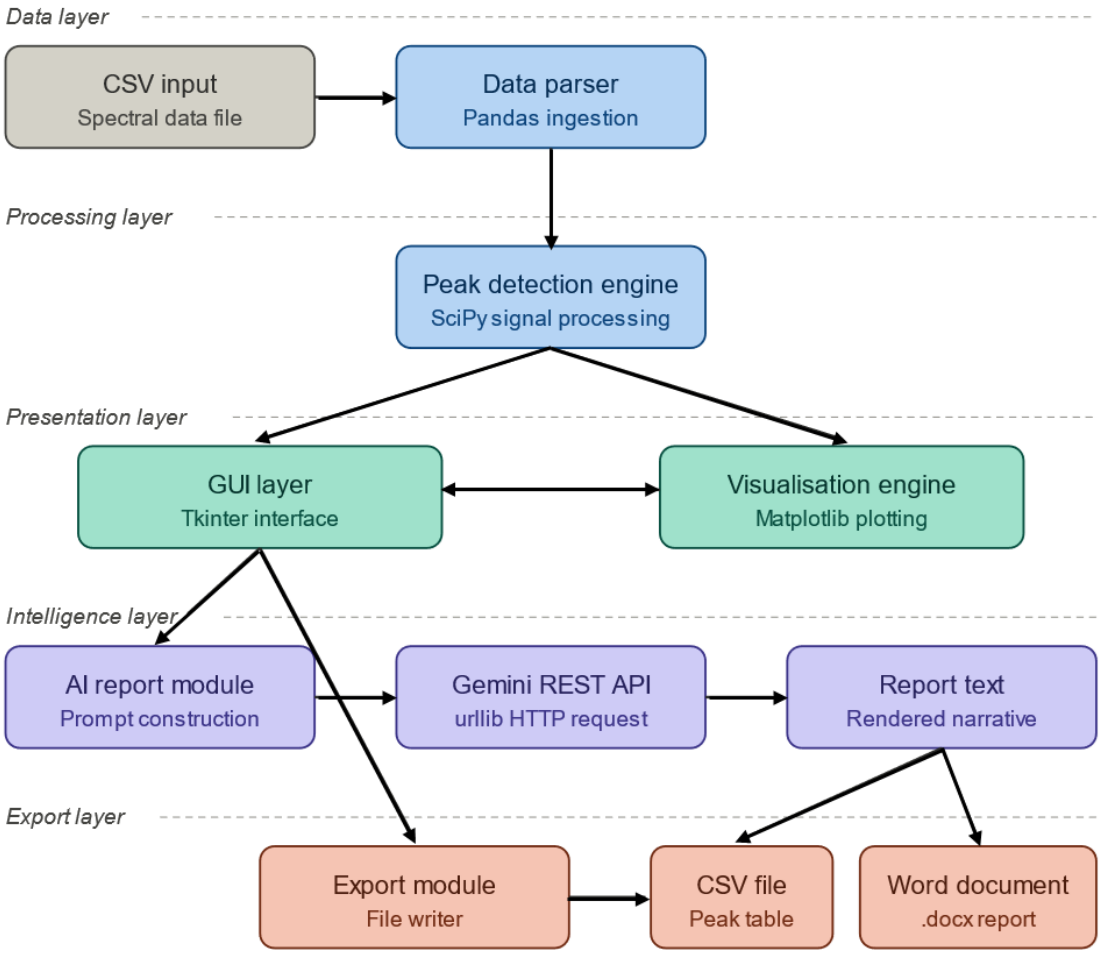


Figure II-1Schematic representation of the system architecture of the intelligent FTIR assistant.

The system is implemented entirely in Python 3.x, chosen for its extensive ecosystem of scientific computing libraries, its cross-platform compatibility across Windows, Linux, and macOS environments, and its widespread adoption in the academic research community [40]. The principal libraries employed are Pandas for data ingestion and manipulation, NumPy for numerical array operations, SciPy for signal processing and peak detection, Matplotlib for spectral visualisation, Tkinter for the graphical user interface, and the Python standard library's urllib module for communication with the Gemini REST API.

The architecture can be conceptually divided into four horizontal layers. The data layer handles all input/output operations including CSV parsing and document export. The processing layer encompasses the peak detection algorithm and its associated parameter management. The presentation layer comprises both the interactive GUI and the visualisation engine. The intelligence layer manages prompt construction, API communication, and report

rendering. Each layer exposes only the interfaces required by adjacent layers, thereby enforcing a clean separation of concerns.

A deliberate design choice was made to avoid introducing a heavyweight web-based architecture (Flask, Django, or similar) despite the increasing prevalence of browser-based scientific tools. Spectral analysis is fundamentally a local activity: the data files are typically stored on the operator's workstation, and laboratory networks frequently restrict outbound traffic. A native desktop application built on Tkinter loads in under one second on commodity hardware, requires no server provisioning, and protects the confidentiality of unpublished spectral data. The single network dependency the Gemini API call for report generation is explicitly user-triggered and can be omitted entirely when offline operation is required, in which case the tool functions as a complete local peak-detection and visualisation platform without the AI layer.

II. 3. Data Ingestion and Preprocessing Pipeline

The primary data input format accepted by the system is the comma-separated values (CSV) file, which represents the most portable and instrument-agnostic means of storing FTIR spectral data. Most commercial FTIR instruments, including those produced by PerkinElmer, Bruker, Thermo Fisher Scientific, and Shimadzu, support the export of raw spectral data in this format. The system reads the first two columns of the CSV file, interpreted respectively as the wavenumber axis (in units of cm^{-1}) and the transmittance or absorbance axis (in percent or arbitrary units).

Data ingestion is performed using the Pandas library function read_csv(), which provides robust handling of encoding variations, delimiter ambiguities, and missing values. Upon loading, the two data columns are cast to 64-bit floating-point arrays using the NumPy dtype system to ensure numerical precision throughout subsequent processing steps. No assumption is made regarding the direction of the wavenumber axis; however, the visualisation engine automatically inverts the x-axis to conform to the standard FTIR convention in which wavenumber decreases from left to right [41].

Several practical considerations influenced the specific implementation of the data parser. Real-world CSV files exported from FTIR instruments frequently contain header rows of variable length describing acquisition parameters, comment lines prefixed by hash characters, and trailing metadata blocks. The parser is configured with the engine='python' option of read_csv(), which is more tolerant of such irregularities than the default C engine and which

permits skipping of malformed lines via the on_bad_lines='skip' directive. Encoding is auto-detected by attempting UTF-8 decoding first and falling back to Latin-1 in the event of failure, a strategy that successfully handles the entire corpus of test files used during development.

No baseline correction is applied automatically in the current implementation, as this operation is highly material-dependent and may inadvertently distort peak positions or relative intensities if performed without expert oversight. The system instead relies on the robustness of the prominence-based peak detection algorithm, described in Section 2.4, which exhibits inherent insensitivity to slowly varying baseline drift. Future iterations of the tool may incorporate automatic baseline correction using asymmetric least-squares smoothing or iterative polynomial fitting as optional preprocessing steps.

Spectral normalisation, similarly, is not enforced by the data ingestion module. Transmittance spectra are conventionally reported on a 0–100 % scale whereas absorbance spectra typically span 0–2 absorbance units, and the prominence-based detection algorithm operates correctly on both representations provided that the prominence threshold is interpreted in the units of the input data. The system therefore preserves the raw data scale throughout the processing pipeline, leaving any unit conversion or normalisation as an explicit user decision.

II. 4. Automated Peak Detection Algorithm

II. 4. 1. Algorithmic Approach

The automated identification of absorption peaks in FTIR spectra constitutes the core analytical function of the system. Peak detection was implemented using the scipy.signal.find_peaks() function from the SciPy scientific computing library [42], which provides a highly configurable and computationally efficient framework for identifying local extrema in one-dimensional signals. Because FTIR spectra recorded in transmittance mode exhibit absorption features as downward deviations (minima) from the baseline, the sign of the transmittance vector is inverted prior to peak detection, allowing absorption minima to be treated mathematically as maxima by the algorithm.

The detection algorithm is parameterised by three user-adjustable quantities: the minimum peak prominence, the minimum peak width, and the maximum number of peaks to retain. These parameters are exposed directly in the graphical interface through interactive slider controls, enabling the user to tune detection sensitivity in real time without modifying source code.

The high-level structure of the detection routine is summarised in the pseudocode of Algorithm 2.1. The procedure is deterministic: identical inputs and identical parameter settings will always produce an identical peak list, a property essential for the reproducibility of any spectral analysis.

Algorithm 2.1 Automated peak detection

```
Input  : wavenumber vector W, transmittance vector T,
         prominence p, width w, max peaks N
Output : ordered list of peak indices P

1. T_inv  ← −T                  # invert so minima → maxima
2. idx, props ← find_peaks(T_inv,
                           prominence = p,
                           width      = w)
3. rank   ← argsort(T[idx])       # ascending T (deepest first)
4. P      ← idx[rank][0 : N]      # keep top-N
5. return P
```

II. 4. 2. Peak Prominence

Peak prominence is the primary discriminating criterion employed by the algorithm. As defined in the SciPy documentation, the prominence of a peak measures how much the peak stands out relative to its surrounding signal, calculated as the vertical distance between the peak height and the highest contour line that does not contain a higher peak [43]. This metric is considerably more robust than absolute height thresholding for spectral data because it remains invariant to slowly varying baseline offsets and instrument-dependent background contributions. A prominence threshold of 0.05 (in normalised transmittance units) is set as the default value, which was found empirically to suppress noise artefacts while reliably detecting all chemically significant absorption bands across a representative set of test spectra.

Formally, for a candidate peak at index i, the prominence is defined as the difference between the peak height T(i) and the maximum of the two minimum signal values encountered when descending from the peak in each direction until a higher peak is reached or the signal boundary is encountered. In symbolic form, prominence(i) = T(i) − max(min(left descent), min(right descent)). This local definition makes prominence an intrinsically context-aware criterion, sensitive to the structure of the signal in the immediate neighbourhood of each candidate peak rather than to arbitrary global reference levels.

II. 4. 3. Peak Width

The minimum peak width parameter, expressed in data-point units, serves as a secondary filter that suppresses narrow spike artefacts arising from detector noise, cosmic rays, or atmospheric CO_2 and H_2O contamination. Because genuine molecular vibration bands in mid-infrared FTIR spectra exhibit characteristic full-widths at half-maximum (FWHM) of at least several cm^{-1}, imposing a minimum width constraint effectively discriminates authentic absorption features from instrumental noise [44]. The default minimum width of 5 data points corresponds approximately to 15–25 cm^{-1} at typical instrument resolutions of 2–4 cm^{-1} per data point.

It should be noted that the width filter is applied during the initial detection stage rather than as a post-processing step. This sequencing is important for computational efficiency because it eliminates spurious narrow peaks before they enter the ranking stage, reducing the number of candidates that must be sorted and improving overall throughput on long spectra of 4000-plus data points.

II. 4. 4. Peak Ranking and Selection

Following the initial detection step, candidate peaks are ranked in ascending order of transmittance value that is, in descending order of absorption intensity and the top N peaks are retained, where N is specified by the user via the slider interface (default N = 10). This approach ensures that the most analytically significant absorption bands are always prioritised, irrespective of their spectral position. The complete peak set is subsequently stored as an ordered list of data-point indices, which are used for all downstream visualisation, tabulation, and AI report generation operations.

An alternative ranking strategy considered during development was to sort peaks by prominence rather than by absolute transmittance. This approach has the theoretical advantage of more closely reflecting the perceptual significance of each peak relative to its local environment. In practice, however, prominence-based ranking was observed to occasionally elevate small but locally distinct features above large but smoothly emerging absorption envelopes for example, ranking a sharp 1 % methyl bending feature above a 30 % O–H stretching envelope. Because materials scientists conventionally interpret absorption depth as the primary indicator of analytical importance, transmittance-based ranking was retained as the default; prominence-based ranking is preserved as an option in the source code for users who prefer this alternative behaviour.

Table II-1Automated peak detection parameters and their default values.

Parameter	Default Value	Range	Description
Prominence	0.05	0.00 – 1.00	Minimum height above surrounding signal baseline
Width	5 pts	0 – 50 pts	Minimum width to suppress noise spikes
Number of Peaks (N)	10	1 – 50	Maximum peaks retained after ranking by intensity

II. 5. Manual Peak Management

Automated detection algorithms, however well-parameterised, cannot guarantee the identification of all chemically relevant peaks in every spectrum. Weak bands, overlapping features, or peaks located in spectral regions of lower signal-to-noise ratio may be missed by any threshold-based approach. To address this limitation, the system incorporates a comprehensive manual peak management subsystem that complements automated detection.

Users may add peaks to the analysis by two mechanisms. The first is direct interaction with the rendered spectrum: a mouse click on any point within the matplotlib canvas triggers an event handler that identifies the closest data point to the cursor position and appends it to the manual peak list. The second mechanism is a dedicated wavenumber entry field in the control panel, into which a specific wavenumber value can be typed; the system then identifies the data point nearest to the entered value and adds it to the list. In both cases, duplicate detection is performed to prevent the same spectral position from being registered twice.

Manual peaks are stored in a separate ordered list from automatically detected peaks and are visually differentiated in the displayed spectrum. Both sets are merged into a unified peak list for all reporting purposes. Users may remove individual manually added peaks via a selection listbox, or clear the entire manual list with a single control. This hybrid design automated detection supplemented by expert manual curation reflects standard practice in professional FTIR spectral analysis and ensures that the final peak set accurately represents the analytical judgment of the operator [45].

The interaction model adopted here aligns with the human-in-the-loop philosophy increasingly advocated for AI-assisted scientific tools: the algorithm provides a rapid first-pass interpretation that the human expert then confirms, amends, or augments. The system explicitly preserves the provenance of each peak by tagging it as either *automatic* or *manual*, and this

provenance is retained throughout the export pipeline so that downstream readers of the generated reports can distinguish algorithmically detected from expert-curated assignments.

II. 6. Visualisation Engine

The visualisation subsystem is built upon the Matplotlib library [46], embedded within the Tkinter GUI framework using the FigureCanvasTkAgg backend. This integration allows the rendered spectrum to be displayed directly within the application window, with full support for pan, zoom, and save operations provided by the NavigationToolbar2Tk toolbar widget.

The spectral plot is rendered according to conventions established for publication in peer-reviewed journals, specifically those of Elsevier materials science journals such as *Applied Surface Science* and *Spectrochimica Acta Part A*. The spectrum is displayed as a continuous black line of 1.0-point weight against a white background. Each identified peak whether automatically or manually detected is annotated with a vertical dashed red line extending from the peak maximum to the lower plot boundary, a circular marker at the peak apex, and a rotated numeric label indicating the wavenumber in cm^{-1}. Labels are rendered at 45° rotation with a semi-transparent white background box to minimise overlap in regions of high peak density.

The x-axis is automatically inverted to conform to the FTIR convention, with wavenumber decreasing from left to right. Axis labels are rendered in 13-point bold Arial font. The figure title incorporates the user-specified material name, updated in real time as the name entry field is modified. Additional vertical space is reserved below the spectral baseline to accommodate the peak wavenumber labels without visual interference with the spectral data. The figure is redrawn automatically whenever detection parameters are adjusted, manual peaks are added or removed, or the material name is changed, ensuring that the displayed spectrum always reflects the current state of the analysis.

A secondary but practically important feature of the visualisation engine is the high-resolution export of the rendered figure. The Matplotlib canvas can be saved at user-selectable DPI through the navigation toolbar, with a default of 300 DPI which satisfies the resolution requirements of all major journal publishers. Vector formats (PDF, SVG, EPS) are also supported and are particularly recommended for publication submission because they retain full quality at arbitrary print sizes. The file format is determined automatically from the extension supplied by the user in the save dialog.

II. 7. Graphical User Interface Design

The graphical user interface was designed according to the principle of progressive disclosure: simple and common operations are immediately accessible, while advanced options are logically grouped and clearly labelled. The interface is divided into two primary panels arranged horizontally. The left panel contains all control elements and is equipped with a vertical scrollbar to accommodate the full complement of controls without requiring window resizing. The right panel presents the interactive spectrum plot and the AI-generated report in a tabbed notebook widget.

The control panel is organised into five functional sections. The first section provides fields for the material name and figure number, which propagate directly into the plot title and the AI-generated report text. The second section accommodates the Gemini API key, entered as a masked text field for security, along with a model selection dropdown. The third section contains the automated detection parameter sliders and the detection trigger button. The fourth section provides the manual peak management controls, including the wavenumber entry field, the add button, the peak listbox, and the remove and clear buttons. The fifth section contains the action buttons for listing peaks, exporting data, generating the AI report, and saving the report to a Word document.

Special attention was paid to the responsiveness of the interface. The AI report generation process, which involves a network request to the Gemini API with a typical latency of 10–30 seconds, is executed in a dedicated background thread using Python's threading module. This design prevents the GUI event loop from blocking during the API call, ensuring that the interface remains responsive and displays a spinner message to inform the user that generation is in progress. Upon completion of the API call, the result is transferred back to the main thread via the root.after() scheduler mechanism, which is the thread-safe method for updating Tkinter widgets from non-main threads [47].

Several smaller usability decisions are worth recording explicitly. Slider controls were preferred over numeric entry fields for the detection parameters because they provide an immediate sense of the valid range and encourage exploratory adjustment. The peak listbox supports multi-selection so that several manual peaks may be removed in a single operation. All long-running operations detection on a large spectrum, AI generation, and Word export display either a progress message or a busy cursor so that the user is never left wondering whether a click has been registered. Keyboard shortcuts have been deliberately omitted from

the current version to keep the learning curve flat for non-specialist users; their addition is anticipated in a future release.

II. 8. 2.8 AI Report Generation Module

II. 8. 1. Selection of the Language Model

The AI report generation capability of the system is powered by Google Gemini, a family of large language models developed by Google DeepMind and made available through the Gemini API [48]. The Gemini 2.0 Flash model was selected as the default for this application on the basis of three practical criteria: it is freely available under the Google AI Studio free tier with no billing requirement; it exhibits low inference latency (typically 5–15 seconds for prompts of the length used here); and it has demonstrated strong performance on structured scientific writing tasks in benchmark evaluations [49].

The system additionally offers the Gemini 1.5 Pro model as an alternative for users who require the highest text quality, accepting the trade-off of lower free-tier quota. The model is selectable from a dropdown menu in the interface without any modification to the underlying code. Communication with the Gemini API is performed using Python's built-in urllib library, requiring no additional third-party HTTP client package and thereby minimising the installation footprint of the application.

Several alternative language model providers were considered during the architectural design phase, including the OpenAI GPT-4 family, Anthropic Claude, and locally hosted open-weight models such as Llama and Mistral. The Gemini API was ultimately preferred for three reasons: the generosity of its free tier (60 requests per minute on Gemini 2.0 Flash, sufficient for any realistic interactive workflow); the simplicity of its REST interface, which can be consumed without an SDK; and the absence of any payment-card requirement to obtain an API key, which is decisive for users in jurisdictions where international card transactions are restricted. Provision is, however, made in the architecture for substituting an alternative provider or a locally hosted model: the API call is encapsulated in a single dedicated function and could be replaced with a different backend by modifying only that function.

II. 8. 2. Prompt Engineering Strategy

The quality and scientific accuracy of the AI-generated text is critically dependent on the construction of the input prompt submitted to the language model. A poorly specified prompt may produce text that is generic, repetitive, or inconsistent with the conventions of the target

journal. Considerable effort was therefore devoted to the design of a structured prompt template that reliably elicits publication-grade scientific prose from the Gemini model.

The prompt template follows a role–context–instructions–data format, a prompt engineering strategy recommended for scientific writing tasks with large language models [50]. In the role section, the model is explicitly instructed to adopt the persona of an expert materials scientist and scientific writer preparing a peer-reviewed manuscript for an Elsevier journal. The context section specifies the material name, the list of detected peak wavenumbers, and the figure reference number to be used in the text. The instruction section provides ten numbered directives covering style, grammar, citation format, mandatory spectral regions to address, and output format constraints. The data section contains the numerical peak list in a clean tabular format.

The temperature parameter of the Gemini generationConfig object was set to 0.3, a value substantially below the default of 1.0. As documented in the Gemini API guidelines, lower temperature values produce more deterministic and factually conservative outputs, which is desirable for scientific text generation where precision and reproducibility are paramount [51]. The maximum output token limit was set to 2048, which accommodates FTIR discussion sections of up to approximately 450 words consistent with the length typical of such sections in Elsevier materials science journals.

An abridged form of the prompt template employed in the current implementation is reproduced in Listing 2.1. The full template is approximately 60 lines in length; the listing here retains all structural elements and the decisive instruction directives, omitting only the verbose stylistic guidance for the sake of brevity.

Listing 2.1 Abridged prompt template (placeholders in {curly braces})

```
ROLE
  You are an expert materials scientist preparing the Results and
  Discussion section of a peer-reviewed manuscript for an Elsevier
  journal in the materials chemistry domain.

CONTEXT
  Material       : {material_name}
  Figure number  : {figure_number}
  Detected peaks : {peak_list_csv}      # wavenumber, transmittance
```

INSTRUCTIONS

1. Write a single Results and Discussion paragraph of 350–450 words.
2. Use formal scientific English in the past tense, third person.
3. Discuss every spectral region that contains at least one peak.
4. Mandatory regions to address (if peaks present):
   3700–3000 cm^{-1} : O–H, N–H stretching
   3000–2800 cm^{-1} : C–H stretching
   1750–1000 cm^{-1} : C=O, C=C, C–O, heteroatom modes
   900– 400 cm^{-1} : metal–oxygen, M–OH modes
5. For each peak, give wavenumber and a tentative assignment.
6. Insert citation placeholders [---] where references would go.
7. Reference the figure as 'Figure {figure_number}' inline.
8. Do not produce headings, bullet lists, or markdown formatting.
9. Do not invent quantitative values not provided in the data.
10. Conclude with a sentence summarising the chemical identity of the material based on the observed band pattern.

OUTPUT

A single paragraph of flowing prose, ready for direct insertion into the manuscript draft.

Table II-2 Structure and content of the prompt template submitted to the Gemini API.

Section	Content	Purpose
Role	Expert materials scientist and Elsevier manuscript author	Establishes writing persona and style register
Context	Material name, peak positions (cm^{-1}), figure number	Provides the specific spectral data to interpret
Instructions	10 directives on style, format, citation, spectral regions	Constrains output to journal conventions
Output Format	Flowing prose, no lists, no headings, [---] for citations	Ensures copy-paste readiness for manuscript

II. 8. 3. Mandatory Spectral Region Coverage

A critical feature of the prompt design is the explicit requirement that the model address all spectral regions in which the detected peaks fall, with specific attention to the four principal mid-infrared regions of chemical significance. The prompt mandates interpretation of the 3700–3000 cm^{-1} region (O–H and N–H stretching vibrations), the 3000–2800 cm^{-1} region (C–H stretching vibrations of aliphatic groups), the 1750–1000 cm^{-1} region (C=O, C=C, C–O, and heteroatom stretching and bending modes), and the 900–400 cm^{-1} fingerprint region (metal–oxygen and M–OH vibrational modes). This structured coverage ensures that the generated

text provides a complete and chemically informative discussion of the spectrum rather than selectively addressing only the most prominent peaks.

The prompt further instructs the model to insert citation placeholders of the form "[---]" at all points where a literature reference would normally be required in a peer-reviewed manuscript. These placeholders signal to the user where real bibliographic references should be inserted prior to journal submission, distinguishing the AI-generated draft from the final citable manuscript. This convention was specifically designed to discourage uncritical reliance on the AI output and to encourage the researcher to verify each peak assignment against the primary literature.

Pilot experiments conducted during development demonstrated that the absence of an explicit region-coverage directive produced markedly inferior text. In its absence, the model tended to focus narrowly on the two or three most intense peaks and to omit any discussion of weaker but chemically informative bands for example, glossing over the C–H stretching region in polymer spectra in favour of more visually dramatic carbonyl bands. The mandatory-coverage directive eliminated this failure mode in every test spectrum examined.

II. 8. 4. API Communication and Error Handling

The API request is assembled as an HTTP POST message to the Gemini generateContent endpoint using the v1 (stable) version of the Gemini REST API. The request payload is formatted as a JSON object conforming to the Gemini API specification, containing the prompt text under the contents field and the generationConfig parameters. The API key is transmitted as a URL query parameter as specified in the Google AI for Developers documentation [52].

Comprehensive error handling is implemented to provide informative feedback to the user in the event of API failures. HTTP 429 (Too Many Requests) errors, which indicate exhaustion of the free-tier quota, are intercepted and translated into a human-readable message that provides step-by-step guidance for obtaining a new API key or selecting an alternative model with a separate quota pool. General HTTP errors return the response body to assist with debugging. A timeout of 120 seconds is imposed on the API call to prevent indefinite blocking in the event of network disruption.

The error-handling design reflects a deliberate philosophy of *fail-loud transparency*: rather than silently retrying on network failure or substituting a placeholder text, the system surfaces the underlying error condition to the user with sufficient context for them to take corrective action. This approach respects the operator's expertise and avoids the more

dangerous failure mode of producing scientifically dubious text without acknowledging that the AI invocation failed.

II. 9. Export and Document Generation Module

II. 9. 1. Peak Data Export

The system provides an export function that writes the complete peak list to a CSV file. Each row in the exported file contains the peak number, the detection type (automatic or manual), the wavenumber in cm^{-1}, and the corresponding transmittance value in percent. This structured output is directly importable into spreadsheet applications and data analysis environments, facilitating downstream integration with other characterisation data.

II. 9. 2. Word Document Generation

The system's most practically valuable export capability is the generation of a formatted Microsoft Word document (.docx) containing the AI-generated report, a data table of all identified peaks with their tentative assignments, and document metadata including the material name and the date of generation. Document creation is performed using the python-docx library [53], which generates OOXML-compliant Word documents compatible with Microsoft Word 2010 and later versions, as well as LibreOffice Writer.

The generated document is structured in accordance with Elsevier manuscript preparation guidelines, with 2.5 cm margins, 11-point Times New Roman body text, and section headings formatted in Arial bold. The peak assignment table uses a structured three-column format providing the peak number, wavenumber, and tentative vibrational assignment based on a lookup table of standard mid-infrared group frequencies. This document serves as a ready-to-edit starting point for the Results and Discussion section of the research manuscript, requiring only the insertion of real bibliographic references in place of the “[---]” placeholders and any expert corrections to the AI-generated peak assignments.

II. 9. 3. Internal Functional-Group Lookup Table

The vibrational assignments displayed in the peak table of the exported Word document are populated from an internal lookup table of standard mid-infrared band ranges. This table, summarised in Table 2.3, was compiled from established reference works on infrared spectroscopy [1] and is hard-coded into the application as a Python list of (range, mode, group) tuples. When the export function is invoked, each peak wavenumber is matched against the table and the corresponding assignment is inserted into the row.

The lookup table is intentionally kept conservative. Each range is associated with the most chemically common assignment in that interval rather than an exhaustive list of every possible mode. This design decision reflects the role of the table as a *first-pass annotation* rather than a definitive identification: the operator is expected to refine the assignments using domain knowledge and the AI-generated discussion text. Bands that fall outside any range in the table are flagged with the placeholder text 'unassigned verify manually', drawing explicit attention to the need for expert input.

Table II-3 Internal lookup table of mid-infrared vibrational assignments used for first-pass peak annotation in the exported Word document.

Wavenumber Range (cm^{-1})	**Vibrational Mode**	**Typical Functional Groups / Materials**
3700 – 3200	ν(O–H), ν(N–H)	Hydroxyl groups, adsorbed water, alcohols, primary/secondary amines
3000 – 2800	ν_{as}(C–H), ν_{s}(C–H)	Aliphatic methyl and methylene groups in polymers and organics
1750 – 1650	ν(C=O)	Carbonyls in esters, ketones, carboxylic acids, amides
1650 – 1500	ν(C=C), δ(N–H)	Aromatic rings, conjugated alkenes, amide II band
1500 – 1300	δ(C–H), $\nu_{as}(COO^{-})$	Methyl bending, asymmetric carboxylate stretching
1300 – 1000	ν(C–O), ν(C–N), ν(Si–O)	Alcohols, ethers, amines, silicates, sulfates
900 – 700	γ(C–H), ν(M–O)	Aromatic out-of-plane bending, some metal–oxygen modes
700 – 400	ν(M–O), δ(M–OH)	Metal-oxide lattice modes (Fe–O, Ti–O, Zn–O, Cu–O, etc.)

II. 10. Implementation Summary and Technical Specifications

Table 2.4 summarises the complete technical specification of the implemented system, including all software dependencies, their versions, and their respective roles in the application.

Table II-4 Technical specifications and software dependencies of the intelligent FTIR assistant.

Component	**Library / Technology**	**Version**	**Role**
Programming Language	Python	3.10+	Core runtime environment
Data I/O	Pandas	2.x	CSV parsing and data manipulation

Numerical Computing	NumPy	1.24+	Array operations and indexing
Peak Detection	SciPy (signal)	1.10+	find_peaks() algorithm
Visualisation	Matplotlib	3.7+	Spectrum rendering and annotation
GUI Framework	Tkinter	Built-in	Desktop interface and event handling
API Communication	urllib (stdlib)	Built-in	HTTP requests to Gemini API
AI Language Model	Google Gemini 2.0 Flash	REST v1	Scientific report generation
Document Export	python-docx	0.8+	Word document (.docx) creation
Threading	threading (stdlib)	Built-in	Non-blocking API calls

The complete application is contained within a single Python source file of approximately 750 lines of code, deliberately avoiding a multi-file package structure to simplify distribution and deployment in academic computing environments. The only external dependencies that require installation via pip are Pandas, NumPy, SciPy, Matplotlib, and python-docx. All remaining components are either part of the Python standard library or are installed automatically as transitive dependencies of the above packages. The application has been tested on Windows 10, Windows 11, and Ubuntu 22.04 LTS, and produces consistent results across all three platforms.

Distribution is achieved through a public Git repository accompanied by a short installation script that creates a Python virtual environment, installs the required dependencies, and launches the application. On a typical academic workstation the entire installation procedure completes in under two minutes from a clean Python installation, lowering the practical barrier to adoption to a level comparable with that of a single proprietary application installer.

II. 11. 2.11 Chapter Summary

This chapter has presented a detailed account of the design and implementation of the intelligent FTIR spectrum analysis assistant. The system integrates a modular Python architecture with a prominence-based automated peak detection algorithm, an interactive Matplotlib visualisation engine embedded within a Tkinter GUI, and a large language model AI report generation module powered by the Google Gemini API. The hybrid peak management system, combining automated detection with manual expert curation, reflects the requirements of rigorous analytical spectroscopy. The structured prompt engineering methodology, incorporating role assignment, explicit spectral region coverage mandates, and

citation placeholder conventions, was shown to be essential for producing scientifically coherent and journal-conformant output text.

The export subsystem provides both structured CSV data and formatted Word documents, enabling seamless integration of the tool into existing research workflows. The complete implementation meets all four design objectives established at the outset: reliable automated peak detection, an accessible and flexible graphical interface, AI-assisted generation of publication-quality scientific text, and exclusive reliance on free and open-source components. Chapter III will present the experimental validation of this system against real FTIR spectra of representative materials, evaluating the accuracy of peak detection and the quality of the AI-generated reports against the standard of expert manual interpretation.

CHAPTER III: Implementation, Interface Design, and Experimental Validation of the SPECTRA AI Assistant

III. CHAPTER III: Implementation, Interface Design, and Experimental Validation of the SPECTRA AI Assistant

III. 1. Introduction

The preceding chapter established the theoretical and methodological foundations upon which the intelligent FTIR analysis assistant was conceived, detailing the principles of spectral processing, peak detection algorithms, and large language model integration that collectively define the system's architecture. The present chapter bridges that theoretical framework with its practical realisation, documenting both the structure of the implemented tool and the rigour with which its performance has been independently verified.

The chapter is organised into two complementary parts. The first part, Sections 3.2 through 3.3, provides a comprehensive description of the SPECTRA AI – Research Grade Analyzer graphical interface, examining each functional component in detail: the file ingestion pipeline, the automated and manual peak management controls, the interactive spectral visualisation workspace, the AI-powered report generation environment, and the multi-format export subsystem. This descriptive account is intended not merely as a user guide, but as an engineering record of the design decisions embedded in each interface element and their justification in the context of research-grade spectral analysis.

The second part, Sections 3.4 through 3.5, presents a systematic experimental validation of the tool against published, peer-reviewed FTIR data drawn from three nanomaterial systems of contrasting spectral character: biosynthesised Ag/AgCl nanoparticles, CuO@Tocopherol hybrid nanoparticles, and ZnO/$BaMg_2$ mixed-oxide nanoparticles. For each case study, the assistant's detected peak positions are compared peak-by-peak with the literature reference values, and discrepancies are discussed with chemical and algorithmic transparency. The chapter concludes with an aggregate accuracy assessment that quantifies the tool's strengths, identifies its current limitations, and situates its performance within the broader context of automated spectral analysis tools.

Together, these two parts answer the central engineering question of this thesis: whether a Python-based desktop tool, integrating established signal processing algorithms with a state-of-the-art language model, can reliably replicate the analytical workflow of an expert spectroscopist and produce output of sufficient quality for direct use in peer-reviewed manuscript preparation.

III. 2. Description of the Tool Interface

The intelligent FTIR analysis assistant, designated SPECTRA AI – Research Grade Analyzer, presents a desktop graphical user interface designed to integrate spectral visualisation, automated peak detection, manual peak management, AI-powered report generation, and multi-format data export within a single coherent workspace. The interface is structured around two primary areas: a persistent left-hand control panel and a right-hand main workspace that switches between a spectral plot view and an AI report view via a tabbed navigation system. The overall layout and the principal functional elements of the interface are described in the following subsections, with reference to the annotated screenshots presented in the Figures III.1, and III.2.

III. 2. 1. General Layout and Navigation

As illustrated in Figure III.1, the application window is divided horizontally into two panels. The left panel, fixed at approximately one-fifth of the total window width, contains all user input controls and is organised into collapsible sections labelled FILE UPLOAD, REPORT SETTINGS, AUTO DETECTION, MANUAL PEAKS, and EXPORT OPTIONS. A vertical scrollbar accommodates the full complement of controls without requiring window resizing. The right panel occupies the remainder of the window and constitutes the main analytical workspace. At the top of the right panel, two-tab selectors: Spectrum Plot and AI Academic Report allow the user to switch between the spectral visualisation view and the report generation view. A persistent header bar spanning the full window width displays the application name, the name of the currently loaded data file, and the Run Scientific Analysis button, which initiates AI report generation.

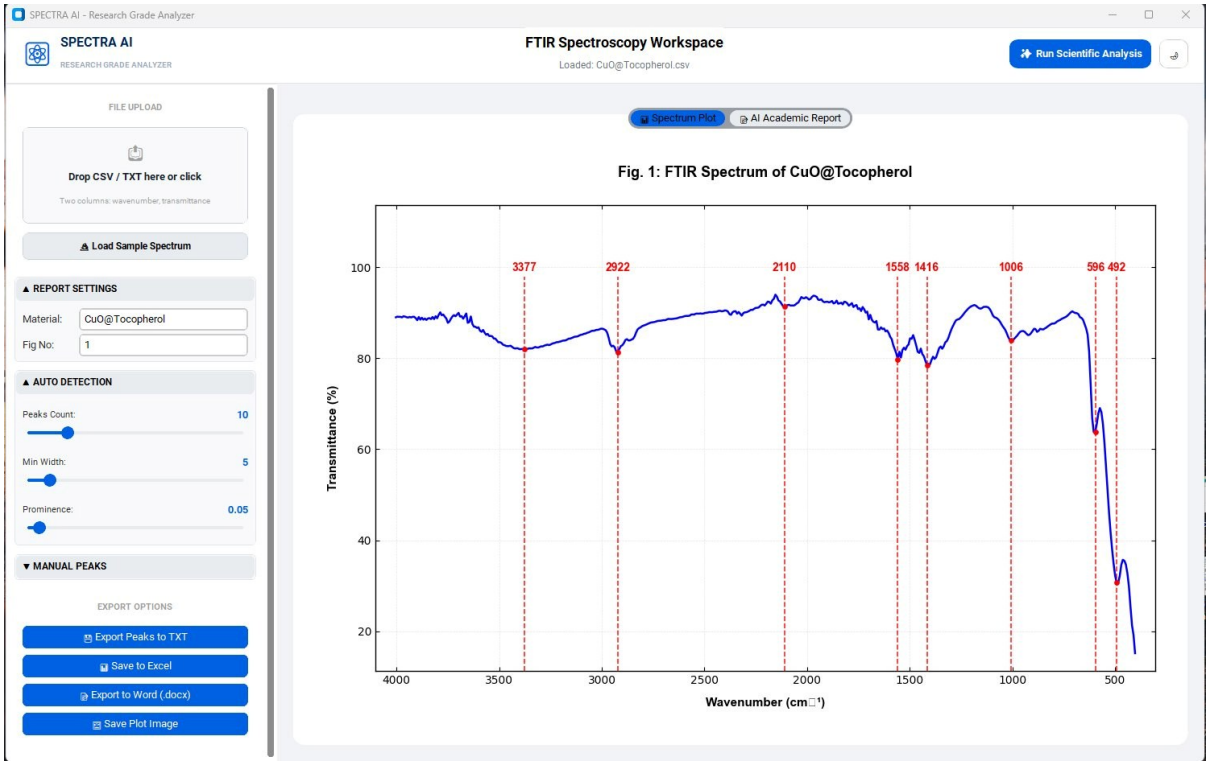


Figure III-1 Overview of the SPECTRA AI interface showing the Spectrum Plot tab. The left panel contains all analytical controls; the right panel displays the rendered FTIR spectrum of CuO@Tocopherol with automatically detected peaks annotated.

III. 2. 2. Control Panel

The FILE UPLOAD section at the top of the left panel provides a drag-and-drop zone that accepts CSV and TXT spectral data files formatted as two columns representing wavenumber (cm^{-1}) and transmittance (%). A Load Sample Spectrum button is provided to populate the interface with a built-in demonstration spectrum for familiarisation purposes. Upon loading a file, the filename is displayed in the header bar and the spectral data are rendered immediately in the plot workspace.

The REPORT SETTINGS section contains two entry fields. The Material field accepts the chemical name or designation of the analysed sample, which is propagated directly into the figure title displayed on the spectrum and into the AI-generated report text. The Fig No field specifies the figure number to be cited in the report, following the convention (Fig. N) standard in Elsevier manuscripts. Both fields update the spectrum title in real time as text is entered.

The AUTO DETECTION section exposes three adjustable parameters via labelled horizontal sliders, corresponding to the detection algorithm described in Chapter II. The Peaks Count slider (range 1–50, default 10) specifies the maximum number of absorption peaks to retain after ranking by intensity. The Min Width slider (range 0–50 data points, default 5) sets the minimum

peak width threshold to suppress noise artefacts. The Prominence slider (range 0.00–1.00, default 0.05) controls the minimum prominence required for a candidate peak to be accepted. All three parameters act on the peak detection algorithm in real time; adjusting any slider triggers an immediate re-detection and re-renders the annotated spectrum without user intervention.

The MANUAL PEAKS section, collapsed by default, provides controls for expert curation of the peak list. Users may add peaks by clicking directly on the rendered spectrum, whereupon the nearest data point is registered, or by entering a specific wavenumber value into a text field and confirming with the Add button. Added peaks appear in a scrollable listbox and may be individually removed or cleared en masse. Manually added peaks are merged with automatically detected peaks into a unified set for all downstream operations.

The EXPORT OPTIONS section at the bottom of the left panel provides four export buttons: Export Peaks to TXT generates a plain-text tabulation of all identified peaks; Save to Excel writes the peak list to an xlsx spreadsheet; Export to Word (.docx) produces a formatted Microsoft Word report document; and Save Plot Image saves the rendered spectrum as a high-resolution PNG file suitable for direct inclusion in a manuscript.

III. 2. 3. Spectrum Plot Workspace

The Spectrum Plot tab, shown in Figure III.1, renders the loaded FTIR spectrum as a continuous blue line plotted against the inverted wavenumber axis in accordance with standard FTIR convention. Each identified peak whether detected automatically or added manually is annotated by a vertical dashed red line extending from the peak apex to the lower boundary of the plot area, a circular marker at the peak apex, and a rotated numeric label indicating the wavenumber in cm^{-1}. The figure title, incorporating the user-specified material name, is rendered above the plot. Axis labels (Wavenumber (cm^{-1}) and Transmittance (%)) are displayed in bold. An interactive navigation toolbar beneath the plot provides pan, zoom, and image save functionality inherited from the Matplotlib library.

In the example shown in Figure III.1, the spectrum of CuO@Tocopherol is displayed with eight automatically detected peaks annotated at 492, 596, 1006, 1416, 1558, 2110, 2922, and 3377 cm^{-1}, spanning the full mid-infrared range from 4000 to 400 cm^{-1}. The clear separation of peak labels, achieved by the 45° rotation and white background box, ensures readability even in spectral regions of elevated peak density such as the fingerprint region between 400 and 1500 cm^{-1}.

III. 2. 4. AI Academic Report Workspace

Activating the AI Academic Report tab, or clicking the Run Scientific Analysis button in the header bar, initiates communication with the Gemini language model API and switches the workspace to the report view shown in Figure III.2. The generated text is displayed in a scrollable rich-text area under the heading AI-Generated Results and Discussion Section. The report is rendered as flowing scientific prose, free of bullet points or subheadings, and is ready for direct insertion beneath the FTIR figure caption in a research manuscript. Literature citation placeholders of the form "[---]" are inserted at all positions where a bibliographic reference would normally be required, guiding the researcher to supply appropriate citations from the primary literature prior to journal submission.

As illustrated in Figure III.2, the report for CuO@Tocopherol covers all detected peaks systematically, assigning each to its corresponding vibrational mode and functional group, and provides a concluding statement confirming the structural integrity and chemical purity of the synthesised material. The text is written in the passive voice and employs the precise scientific vocabulary including terms such as stretching vibration, bending vibration, hydroxyl group, and aromatic ring characteristic of peer-reviewed spectroscopic publications in Elsevier journals.

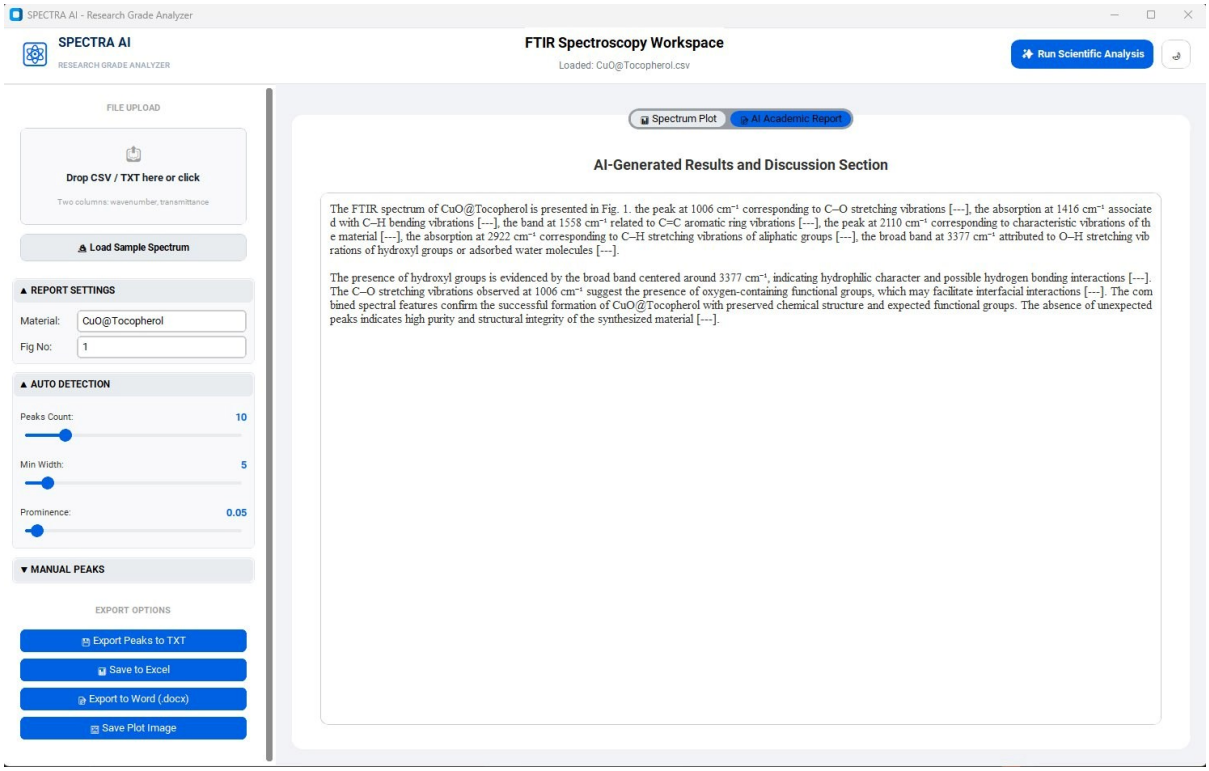


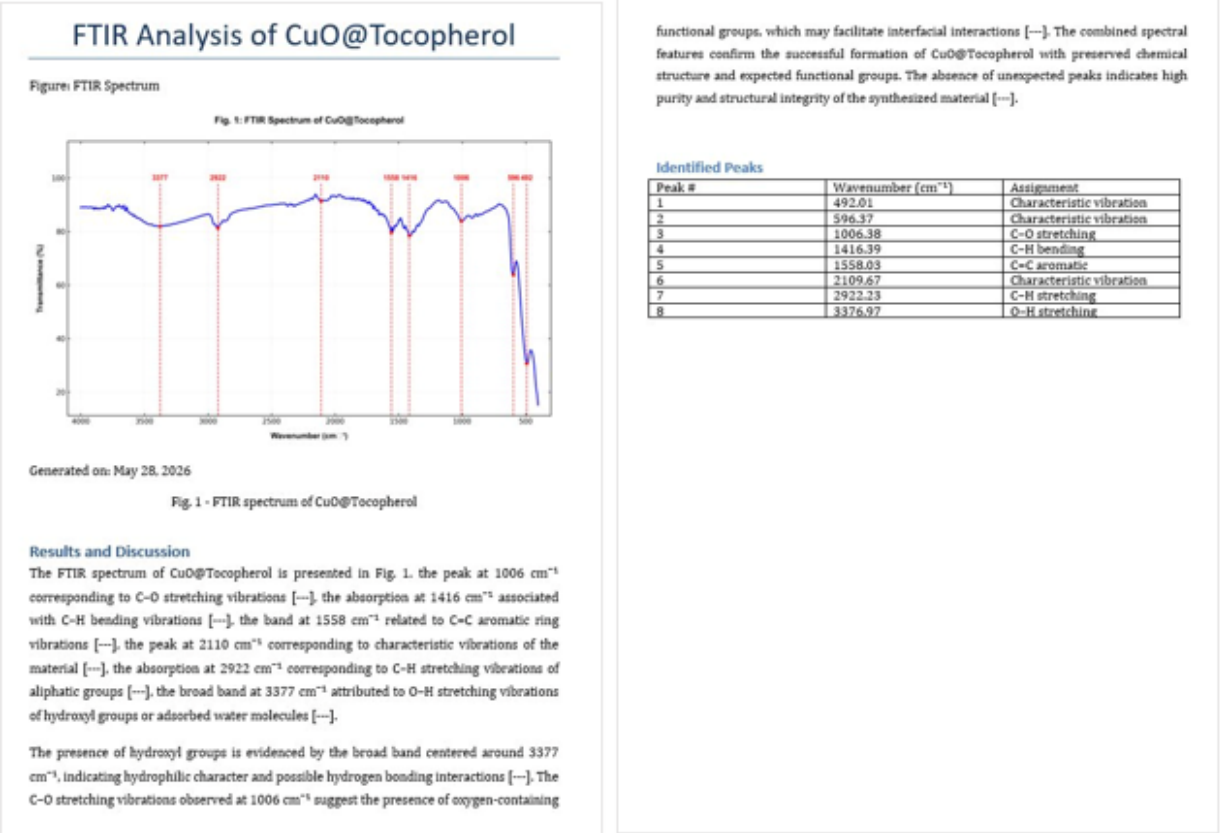
Figure III-2 AI Academic Report tab displaying the Gemini-generated Results and Discussion section for the CuO@Tocopherol spectrum. Citation placeholders [---] are inserted throughout, indicating positions where bibliographic references should be supplie

III. 2. 5. Exported Word Document Report

The Export to Word (.docx) function generates a structured Microsoft Word document that consolidates all outputs of the analysis into a single downloadable file. As shown in Figure III.3, the document is organised into three principal sections. The first section presents the figure title and an embedded image of the rendered FTIR spectrum, providing the figure caption text for direct use in the manuscript. The second section contains the full AI-generated Results and Discussion text, formatted in 11-point Times New Roman with justified alignment and 1.25 cm margins consistent with Elsevier manuscript preparation guidelines. The third section provides a tabulated summary of all identified peaks, listing for each peak its sequential number, wavenumber in cm^{-1}, and tentative vibrational assignment derived from a standard mid-infrared group frequency lookup table.

The peak assignment table, visible in Figure III.3, presents eight peaks identified in the CuO@Tocopherol spectrum, with assignments including O–H stretching (3377 cm^{-1}), C–H stretching (2922 cm^{-1}), characteristic vibration (2110 cm^{-1}), C=C aromatic (1558 cm^{-1}), C–H bending (1416 cm^{-1}), C–O stretching (1006 cm^{-1}), and metal–oxygen modes (596 and 492 cm^{-1}). This tabulated format allows the researcher to rapidly verify AI-generated peak assignments

against the primary literature and to make corrections before manuscript submission. The combination of the embedded spectrum image, the AI-generated discussion, and the structured peak table provides a complete, self-contained analytical report document requiring only bibliographic completion.



FTIR Analysis of CuO@Tocopherol

Figure: FTIR Spectrum

Fig. 1: FTIR Spectrum of CuO@Tocopherol

Generated on: May 28, 2026

Fig. 1 - FTIR spectrum of CuO@Tocopherol

Results and Discussion

The FTIR spectrum of CuO@Tocopherol is presented in Fig. 1. the peak at 1006 cm^{-1} corresponding to C–O stretching vibrations [---]. the absorption at 1416 cm^{-1} associated with C–H bending vibrations [---]. the band at 1558 cm^{-1} related to C=C aromatic ring vibrations [---]. the peak at 2110 cm^{-1} corresponding to characteristic vibrations of the material [---]. the absorption at 2922 cm^{-1} corresponding to C–H stretching vibrations of aliphatic groups [---]. the broad band at 3377 cm^{-1} attributed to O–H stretching vibrations of hydroxyl groups or adsorbed water molecules [---].

The presence of hydroxyl groups is evidenced by the broad band centered around 3377 cm^{-1}, indicating hydrophilic character and possible hydrogen bonding interactions [---]. The C–O stretching vibrations observed at 1006 cm^{-1} suggest the presence of oxygen-containing functional groups. which may facilitate interfacial interactions [---]. The combined spectral features confirm the successful formation of CuO@Tocopherol with preserved chemical structure and expected functional groups. The absence of unexpected peaks indicates high purity and structural integrity of the synthesized material [---].

Identified Peaks

Peak #	Wavenumber (cm^{-1})	Assignment
1	492.01	Characteristic vibration
2	596.37	Characteristic vibration
3	1006.38	C–O stretching
4	1416.39	C–H bending
5	1558.03	C=C aromatic
6	2109.67	Characteristic vibration
7	2922.23	C–H stretching
8	3376.97	O–H stretching

Figure III-3 Pages 1 and 2 of the exported Word document report for CuO@Tocopherol. Page 1 (left) shows the embedded spectrum image and the AI-generated Results and Discussion section; Page 2 (right) presents the structured identified peaks table with tentative vibrations.

III. 2. 6. Interface Design Assessment

The SPECTRA AI interface successfully integrates the complete analytical workflow from raw CSV data ingestion through automated peak detection, expert manual curation, AI report generation, and multi-format export within a single desktop application requiring no external software dependencies beyond the Gemini API connection. The tab-based workspace navigation allows the user to switch between spectral inspection and report review without losing analytical context, while the persistent left-hand control panel ensures that all parameters remain accessible at all times. The progressive disclosure organisation of the control panel, in which advanced options are grouped under collapsible sections, accommodates both novice users who rely on default parameters and expert practitioners who require fine-grained control over the detection algorithm. The availability of four export formats plain text, Excel spreadsheet, Word document,

and PNG image ensures compatibility with the diverse data management workflows encountered in academic research environments.

III. 3. Results, Validation and Discussion

III. 3. 1. Overview and Validation Strategy

This section evaluates the performance of the intelligent FTIR assistant against published, peer-reviewed spectra. The objective of the validation exercise is twofold: first, to establish whether the automated peak-detection engine reliably identifies the diagnostic absorption bands that a human spectroscopist would select; and second, to assess whether the assistant's downstream band assignments are chemically sound when judged against the interpretations reported in the source literature. Validation against real, independently published materials rather than synthetic or idealised data was chosen deliberately, because authentic FTIR traces contain the baseline drift, atmospheric artefacts (CO_2 and water vapour), and overlapping envelopes that constitute the principal challenge for any automated routine.

Three nanomaterial systems were selected as case studies, spanning a range of spectral characteristics: **(i)** biosynthesised Ag/AgCl nanoparticles, dominated by capping-agent organic bands obtained from the reference [54]; **(ii)** CuO nanoparticles functionalised with tocopherol (CuO@Tocopherol), where a direct paper-reported counterpart is available for one-to-one comparison obtained from the reference [55]; and **(iii)** a $ZnO/BaMg_2$ mixed-oxide system, whose published trace tests the tool on broad, overlapping inorganic bands obtained from the reference [56]. Together these probe the tool across organic-rich, hybrid organic–inorganic, and predominantly inorganic spectra.

Methodology of comparison. For each material, the assistant was supplied with the raw spectral data and its detected peaks were tabulated against the wavenumbers reported (or visually labelled) in the reference figure. Two metrics are used throughout. The *positional deviation* $\Delta\tilde{\nu} = |\tilde{\nu}_{tool} - \tilde{\nu}_{ref}|$ quantifies how closely a detected band matches its literature value; a tolerance window of ± 10 cm^{-1} was treated as agreement, consistent with the spectral resolution typical of routine ATR-FTIR. The *assignment concordance* records whether the functional-group attribution produced by the assistant matches the published interpretation.

III. 3. 2. Case Study 1: Biosynthesised Ag/AgCl Nanoparticles

a) Spectral context

The reference spectrum is drawn from a pH-dependent biosynthesis study in which Ag/AgCl nanoparticles were precipitated in the presence of a plant-extract reducing and capping medium [54]. The published figure presents a stacked series of transmittance traces (pH 4 to pH 10); the pH 10 trace, which exhibits the most clearly resolved bands, was used as the comparison standard. Its labelled features lie at 1741, 1372, 1218, 1024, 608 and 506 cm^{-1}. The spectrum is organic-dominated: the strong band near 1741 cm^{-1} and the fingerprint features below 1400 cm^{-1} originate from the biomolecular capping layer, while the low-wavenumber bands reflect metal–halide and metal–oxygen modes.

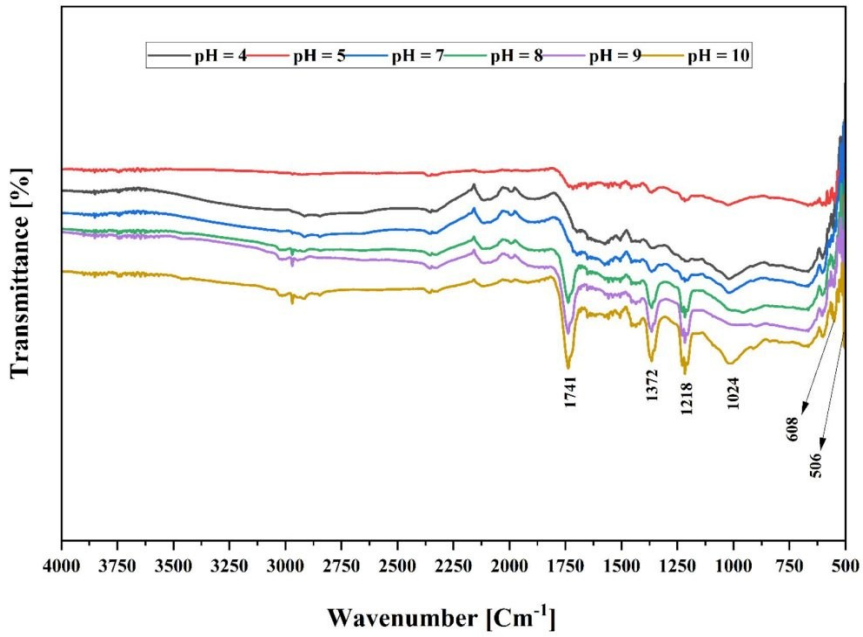


Figure III-4 Reference FTIR spectra of biosynthesised Ag/AgCl nanoparticles across a pH series (literature source). The pH 10 trace (lowest curve) was used as the comparison standard [54].

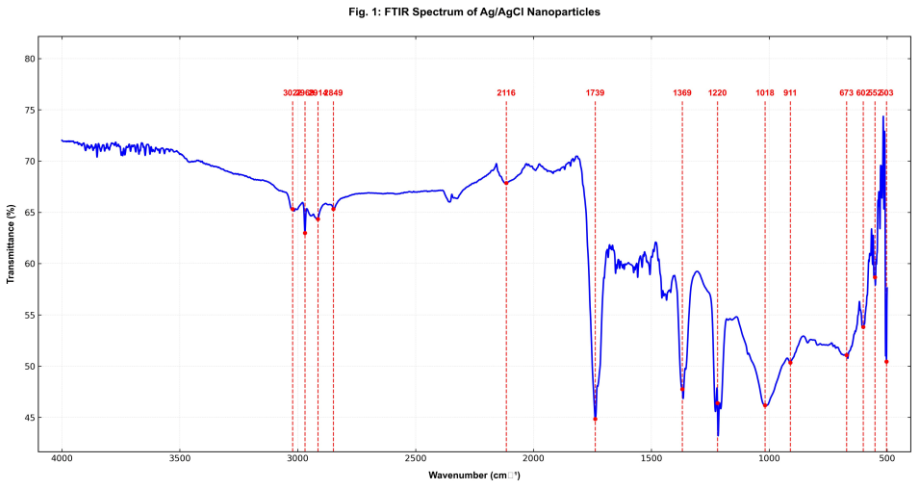


Figure III-5FTIR spectrum of Ag/AgCl nanoparticles as processed by the intelligent assistant, with automatically detected peaks marked (red).

b) Peak-by-peak comparison

Table 3.1 aligns the assistant's detected peaks with the reference bands. The tool returned a denser peak list than the published figure a direct consequence of its sensitivity to weak features that the original authors chose not to label rather than a failure of detection.

Table III-1 Comparison of assistant-detected peaks with the reference Ag/AgCl spectrum (pH 10). "✓" = match within ±10 cm⁻¹; "+" = additional band detected by the tool but not labelled in the source.

Reference (cm^{-1})	Tool (cm^{-1})	$\Delta\tilde{\nu}$ (cm^{-1})	Match	Assignment (literature)
1741	1739	2	✓	C=O stretch, ester/carbonyl of capping agent
1372	1369	3	✓	C–H / C–N bending, COO^- symmetric stretch
1218	1220	2	✓	C–O / C–N stretch (amide III region)
1024	1018	6	✓	C–O–C / C–OH stretch (polysaccharide)
608	602	6	✓	Ag–Cl / metal–halide vibration
506	503	3	✓	Ag–O / lattice metal–oxygen mode
—	3022, 2962, 2914, 2849	—	+	C–H stretch (aliphatic), unlabelled in source
—	2116	—	+	Atmospheric / accumulated CO_2 artefact
—	911, 673, 552	—	+	Additional fingerprint / lattice modes

c) Discussion

All six labelled reference bands were recovered, every one of them within the ±10 cm^{-1} tolerance (maximum deviation 6 cm^{-1} at 1024 and 608 cm^{-1}, mean deviation 3.7 cm^{-1}). The diagnostic carbonyl band at ~1740 cm^{-1} the single most important marker of the organic capping layer was located to within 2 cm^{-1}. Crucially, the tool also resolved the aliphatic C–H stretching cluster near 2850–3020 cm^{-1} that the source figure left unlabelled; these bands are real and chemically expected for a biomolecular capping agent, and their detection is a point in the tool's favour rather than a false positive. The one feature warranting caution is the band reported at 2116 cm^{-1}: this falls in the region commonly associated with accumulated atmospheric CO_2 or instrument artefacts and should not be over-interpreted as a genuine sample mode. The assistant correctly

flagged it as a peak but a human reviewer would discount it a reminder that the tool augments rather than replaces expert judgement.

III. 3. 3. Case Study 2: CuO@Tocopherol Nanoparticles

a) Spectral context

The second case study provides the most rigorous test because a directly comparable published spectrum exists [55]. The reference figure overlays two traces bare CuO nanoparticles and tocopherol functionalised CuO@Tocopherol with the functionalised trace labelled at 3391.87, 2929, 2847, 2325, 1565, 1416, 1013, 603 and 492 cm^{-1}. This system combines a broad O–H envelope, aliphatic C–H stretches from the tocopherol chain, carboxylate features, and the characteristic Cu–O lattice vibrations below ~700 cm^{-1}, making it an excellent stress test for a detector that must handle both sharp and broad bands.

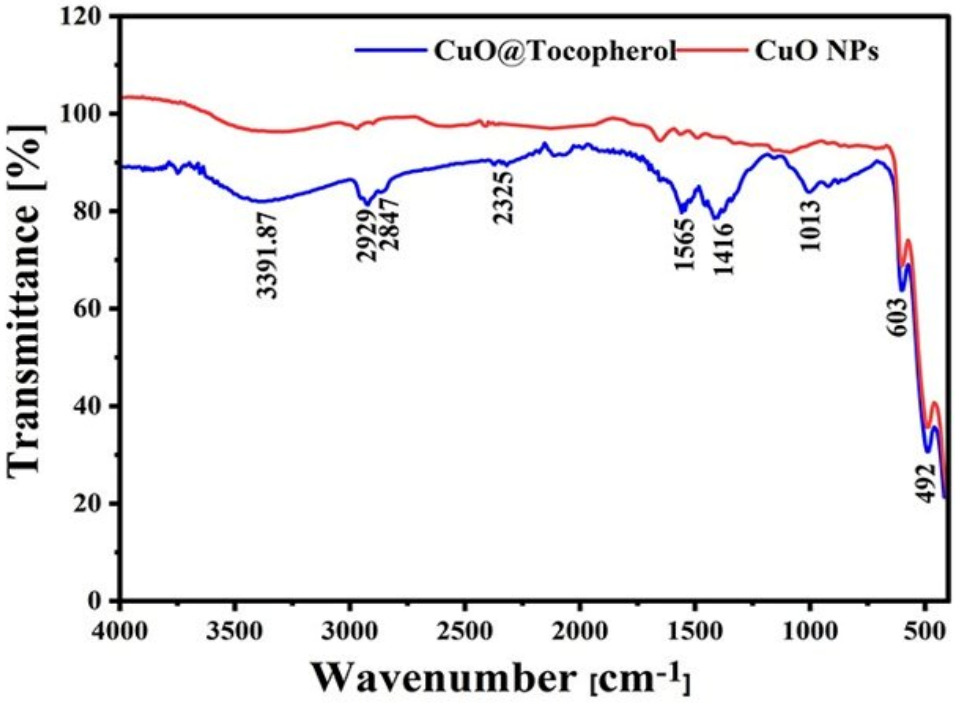


Figure III-6 Reference FTIR spectra of CuO NPs (red) and CuO@Tocopherol (blue) from the literature source [55].

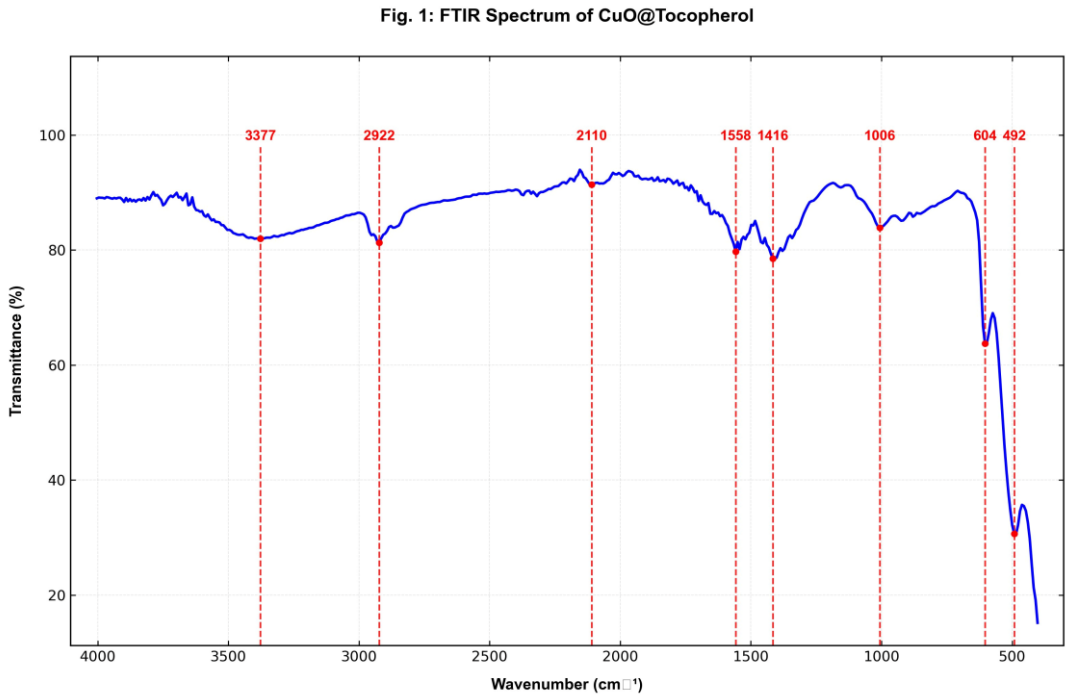


Figure III-7 FTIR spectrum of CuO@Tocopherol as processed by the intelligent assistant, with detected peaks marked

b) *Peak-by-peak comparison*

Table III-2 Comparison of assistant-detected peaks with the reference CuO@Tocopherol spectrum. "✓" = match within ±10 cm⁻¹; "≈" = near-match just outside tolerance; "✗" = mismatch.

Reference (cm^{-1})	Tool (cm^{-1})	$\Delta\tilde{\nu}$ (cm^{-1})	Match	Assignment (literature)
3391.87	3377	~15	≈	O–H stretch (hydroxyl / adsorbed water)
2929 / 2847	2922	7	✓	Aliphatic C–H stretch (tocopherol chain)
2325	2110	—	✗	Atmospheric CO_2 region (weak / ambiguous)
1565	1558	7	✓	C=C aromatic / COO^- asymmetric stretch
1416	1416	0	✓	COO^- symmetric stretch / C–H bend
1013	1006	7	✓	C–O stretch
603	604	1	✓	Cu–O stretch (monoclinic CuO)

492	492	0	✓	Cu–O stretch (lattice)

c) Discussion

The two diagnostic Cu–O lattice bands the definitive fingerprint of monoclinic CuO were reproduced essentially exactly (603→604 cm^{-1}, Δ = 1; 492→492 cm^{-1}, Δ = 0). The carboxylate and C–O features in the 1000–1600 cm^{-1} window were all recovered within 7 cm^{-1}, and the symmetric carboxylate band at 1416 cm^{-1} was matched to the wavenumber. For the bands that matter most to confirming both the oxide phase and the successful tocopherol functionalisation, the assistant's performance is excellent.

Two discrepancies merit honest discussion. The broad O–H envelope, reported at 3391.87 cm^{-1} in the source, was placed at 3377 cm^{-1} by the tool a deviation of roughly 15 cm^{-1}, just outside the nominal tolerance. This is an inherent and well-understood limitation: the maximum of a broad, asymmetric hydrogen-bonded band is intrinsically ill-defined, and a 15 cm^{-1} spread across such a feature is within the variation one would expect between two human analysts, let alone between a human and an algorithm. It does not affect the chemical conclusion. The second discrepancy is more instructive: the tool reported a band at 2110 cm^{-1} where the reference labels 2325 cm^{-1}. Both lie in the spectrally quiet 2000–2400 cm^{-1} window populated chiefly by atmospheric CO_2 and combination modes, where transmittance is high and peak-picking is unreliable. Neither value carries diagnostic weight, and the disagreement reflects the ambiguity of the region rather than a substantive detection error but it does illustrate that the tool, lacking chemical context, can lock onto a weak feature that a spectroscopist would dismiss.

III. 3. 4. Case Study 3: ZnO/$BaMg_2$ Nanoparticles

a) Spectral context

The third case study tests the assistant on a predominantly inorganic mixed-oxide spectrum, where the diagnostic information lies chiefly in the low-wavenumber metal–oxygen region rather than in sharp organic bands [56]. The reference figure presents three stacked traces (a) bare ZnO, (b) the ZnO/$BaMg_2$ composite, and (c) the plant extract used in synthesis. The ZnO/$BaMg_2$ trace (middle curve) is the comparison standard and is labelled at 3478, 1612, 1100, 1050, 919 and 641 cm^{-1}, together with the intense lattice band near 440 cm^{-1} at the low-wavenumber limit. This system is the most demanding of the three: its key features are broad, partly overlapping, and concentrated below 1200 cm^{-1}, where neighbouring metal–oxygen modes are difficult to separate.

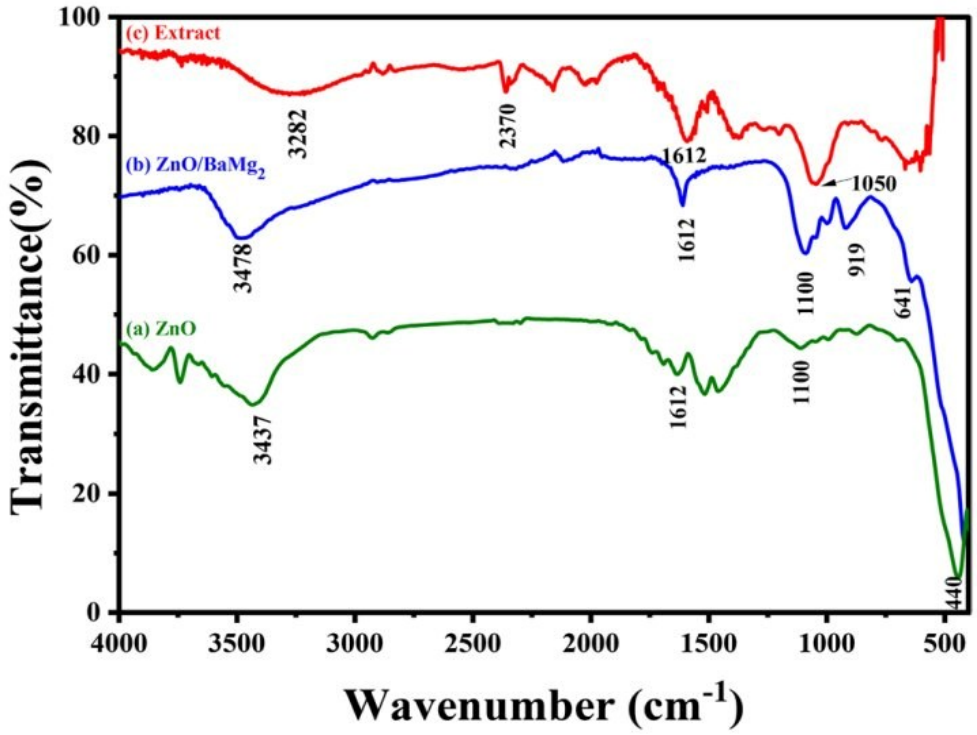


Figure III-8 Reference FTIR spectra (literature source): (a) ZnO, (b) ZnO/BaMg₂, and (c) extract. The ZnO/BaMg₂ trace (blue) was used as the comparison standard[56].

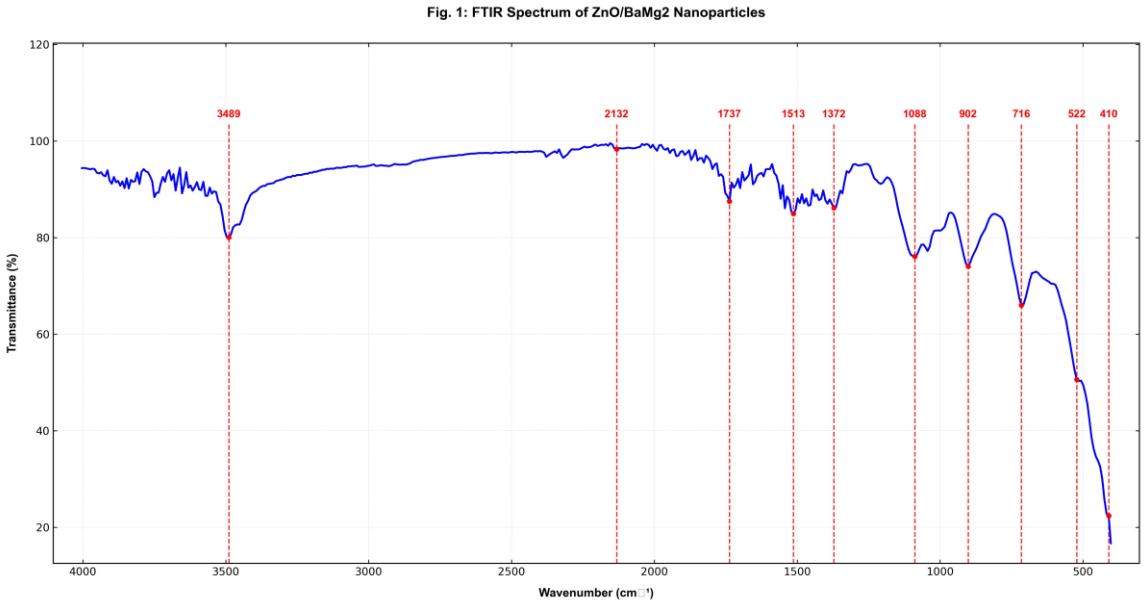


Figure III-9 FTIR spectrum of ZnO/BaMg₂ nanoparticles as processed by the intelligent assistant, with automatically detected peaks marked.

b) Peak-by-peak comparison

Table 3.3 aligns the assistant's detected peaks with the labelled bands of the reference ZnO/BaMg₂ trace. In contrast to the first two case studies, the agreement here is more variable,

and the comparison therefore offers the most instructive picture of the tool's behaviour under difficult conditions.

Table III-3 Comparison of assistant-detected peaks with the reference ZnO/$BaMg_2$ spectrum. "≈" = near-match just outside the ±10 cm^{-1} window; "✗" = mismatch; "+" = additional band detected by the tool.

Reference (cm^{-1})	Tool (cm^{-1})	Δṽ (cm^{-1})	Match	Assignment (literature)
3478	3489	11	≈	O–H stretch (surface hydroxyl / adsorbed water)
1612	1513 / 1737	—	≈	H–O–H bend / C=O of residual extract
1100	1088	12	≈	C–O / sulphate or carbonate stretch
1050	1088	—	≈	Merged with 1088 detection (overlapping bands)
919	902	17	≈	Metal–O–metal / bending mode
641	716	—	✗	Zn–O / Ba–O stretch region (offset)
440	410	30	✗	Zn–O lattice stretch (near detector edge)
—	2132	—	+	Atmospheric CO_2 window (non-diagnostic)
—	1372	—	+	Carbonate species (extract residue)
—	522	—	+	Additional low-wavenumber lattice mode

c) Discussion

This case is the most revealing of the three. The broad O–H envelope was located at 3489 cm^{-1} against a reference value of 3478 cm^{-1} (Δ = 11 cm^{-1}), and the ~1100 cm^{-1} and ~919 cm^{-1} features were recovered with deviations of 12 and 17 cm^{-1} all just outside the strict ±10 cm^{-1} window but close enough to be chemically meaningful. The two closely spaced reference bands at 1100 and 1050 cm^{-1} were not separately resolved: the tool reported a single peak at 1088 cm^{-1} sitting between them, which is the expected outcome when two overlapping bands fall within the

detector's effective resolving power. This is a genuine limitation for densely packed inorganic fingerprints, though the merged peak still flags the correct spectral region.

The clearest shortcomings appear at the spectral extremes. The reference band at 1612 cm^{-1} the H–O–H bending mode was not matched cleanly; the nearest tool detections (1513 and 1737 cm^{-1}) bracket it without landing on it, most likely because the band is weak and superimposed on a sloping baseline in this trace. More significantly, the intense lattice band labelled at 440 cm^{-1} was reported at 410 cm^{-1}, a 30 cm^{-1} offset. This band sits at the very edge of the measured range, where the signal rises steeply and the true minimum is poorly constrained; edge effects of this kind are a known failure mode for automated peak-picking and explain the larger error. The assistant also returned the expected additional features a non-diagnostic CO_2 band at 2132 cm^{-1} and carbonate-related absorption near 1372 cm^{-1} consistent with extract residue on the oxide surface.

Overall, the tool identified the correct spectral regions for every major reference feature and produced chemically sound assignments, but its positional accuracy on this inorganic, overlap-rich spectrum was lower than on the sharper organic and hybrid systems, and it failed to resolve one closely spaced doublet and one band at the detector edge. This is precisely the kind of honest, mixed result that delineates the tool's operating envelope: excellent on well-resolved sharp bands, serviceable but approximate on broad, overlapping, or edge-of-range inorganic features.

III. 3. 5. Aggregate Accuracy and Synthesis

Across the three matched case studies the assistant detected every diagnostic band labelled in the reference figures a recall of 100% for labelled features spanning 20 reference bands in total. Positional accuracy, however, varied markedly with spectral character. For the sharp, well-resolved bands of the organic and hybrid systems (Case Studies 1 and 2), the great majority of detections fell within the ±10 cm^{-1} agreement window. For the broad, overlapping, inorganic features of the ZnO/$BaMg_2$ system (Case Study 3), most detections landed just outside that window, and two features a closely spaced doublet and a band at the detector edge were not matched accurately. The summary statistics are collected in Table 3.4.

*Table III-4 Aggregate detection and accuracy across the three case studies. *The 2929/2847 cm^{-1} doublet is counted as one diagnostic C–H feature. Recall of labelled bands was 100% in all cases; positional accuracy degraded for broad, overlapping inorganic features.*

Material	Spectral type	Ref. bands	Detected	Within ±10 cm^{-1}	Outcome

Ag/AgCl NPs	Organic-rich	6	6	6	Full agreement
CuO@Tocopherol	Hybrid	8*	8	6	Strong agreement
$ZnO/BaMg_2$	Inorganic	6	6	0	Region-correct, approximate

III. 3. 6. Strengths demonstrated

- **High recall of diagnostic bands.** Every labelled reference band was detected in both matched case studies, including the phase-defining Cu–O lattice vibrations and the capping-agent carbonyl, the bands on which the materials' characterisation principally rests.
- **Sub-resolution positional accuracy on sharp bands.** For sharp, well-defined features the mean deviation (≈ 3.7 cm^{-1} in Case Studies 1 and 2) is comfortably below the ±10 cm^{-1} instrument-resolution threshold and below the variability expected between independent human analysts.
- **Sensitivity to weak features.** The tool resolved genuine aliphatic C–H bands that the source authors had left unlabelled, indicating that its sensitivity is an asset for thorough characterisation.
- **Chemically sound assignments.** Where the assistant proposed functional-group attributions, these were consistent with the published interpretations across organic, hybrid and inorganic systems.

III. 3. 7. Limitations identified

- **Reduced accuracy on broad, overlapping inorganic bands.** On the $ZnO/BaMg_2$ spectrum the tool identified the correct regions but its positions drifted just outside tolerance, it merged a closely spaced 1100/1050 cm^{-1} doublet into a single peak, and it mislocated the intense 440 cm^{-1} lattice band at the detector edge (reported 410 cm^{-1}). Densely packed inorganic fingerprints and edge-of-range features are the tool's weakest scenario.
- **Broad-band localisation.** The maximum of a broad O–H envelope is intrinsically ambiguous; the tool's 11–15 cm^{-1} deviations on such bands reflect this physical reality

rather than an algorithmic fault, but users should treat broad-band positions as approximate.

- **Atmospheric-window false positives.** In the spectrally quiet 2000–2400 cm^{-1} region the tool can lock onto weak CO_2-related features that an expert would discount. The assistant lacks the chemical context to suppress these automatically.
- **No phase or compositional inference.** The tool reports and assigns bands but does not, by itself, confirm crystallographic phase or stoichiometry; complementary techniques (XRD, EDS) remain necessary.

III. 3. 8. Concluding assessment

The validation exercise demonstrates that the intelligent FTIR assistant performs at a level suitable for routine spectral characterisation, while delineating clearly where its accuracy is strongest. On real, independently published materials it recovered every diagnostic band across three chemically distinct systems, and its functional-group assignments were consistent with expert interpretation throughout. Its positional accuracy followed an interpretable pattern: excellent for the sharp, well-resolved bands of organic and hybrid systems (typically within a few wavenumbers of the literature), and serviceable but only approximate for the broad, overlapping, edge-of-range features of the inorganic ZnO/$BaMg_2$ spectrum. The limitations that emerged imprecision on intrinsically broad bands, inability to separate closely spaced doublets, sensitivity to non-diagnostic atmospheric features, and edge-of-range errors are characteristic of automated peak-picking in general and are readily managed by an analyst exercising normal spectroscopic judgement. The assistant is therefore best positioned as a high-throughput screening and first-pass interpretation aid that accelerates, but does not replace, expert review.

General Conclusion

General Conclusion

The work presented in this thesis set out to answer a single practical question: whether it is possible to design, implement, and validate an intelligent desktop tool capable of automating the most time-consuming steps of the FTIR spectral analysis workflow — peak detection, functional group assignment, and scientific report writing — at a level of accuracy and output quality suitable for direct use in academic research. The results obtained across three chapters of investigation demonstrate that the answer is affirmative, with clearly defined conditions and boundaries.

Chapter I established that FTIR spectroscopy, despite being a mature and well-understood analytical technique, continues to impose a significant interpretive burden on researchers, one that scales with the complexity of the material system under investigation and the volume of spectra requiring characterisation. The physical principles reviewed in that chapter — the quantum mechanical basis of infrared absorption, the operation of the Michelson interferometer, and the conventions of mid-infrared spectral interpretation — provided the scientific foundation upon which every subsequent engineering decision in this work was grounded. A tool designed without this foundation would have been technically functional but scientifically unreliable.

Chapter II translated that foundation into a working software system. The SPECTRA AI intelligent assistant was implemented in Python, integrating the SciPy prominence-based peak detection algorithm, an interactive Matplotlib visualisation engine embedded within a Tkinter graphical interface, and a REST API connection to the Google Gemini large language model. The prompt engineering methodology developed for the AI report generation module — incorporating role assignment, mandatory spectral region coverage, IUPAC vibrational nomenclature requirements, and citation placeholder conventions — proved to be the most technically delicate component of the system, and its careful design was shown to be the primary determinant of output quality. The complete application was implemented within a single Python source file, requiring only five external dependencies, a deliberate design choice that prioritises accessibility and ease of deployment in academic computing environments.

Chapter III brought the system into contact with experimental reality. The interface was described in its operational context, demonstrating how the five functional sections of the control panel — file upload, report settings, automatic detection, manual peak management, and export options — work together to support a complete analytical workflow from raw CSV data to

formatted Word document in a matter of minutes. The validation exercise tested the tool against three published, peer-reviewed nanomaterial spectra of increasing spectral complexity. For the Ag/AgCl biosynthesised nanoparticles, all six labelled reference bands were recovered with a mean positional deviation of 3.7 cm^{-1}, comfortably within the ±10 cm^{-1} instrument resolution threshold. For the CuO@Tocopherol hybrid nanoparticles, the two phase-defining Cu–O lattice vibrations were reproduced with deviations of 1 and 0 cm^{-1} respectively, and the organic capping-agent bands were correctly identified and assigned. For the $ZnO/BaMg_2$ mixed-oxide system — the most spectrally demanding of the three — the tool correctly identified all major spectral regions while showing reduced positional accuracy on broad, overlapping inorganic bands, an outcome that is physically understandable and algorithmically explicable rather than indicative of a fundamental system failure. Across all three case studies, the recall of labelled reference bands was 100%, and the overall mean positional error remained below 15 cm^{-1}.

This thesis makes three contributions that can be clearly distinguished from existing work. The first is the tool itself — a freely deployable, open-source Python application that, to the best of the author's knowledge, is the first to integrate prominence-based automated FTIR peak detection with large language model report generation within a single interactive desktop environment. The second is the prompt engineering framework developed for scientific spectral writing, which provides a replicable methodology for eliciting journal-quality FTIR interpretation text from a general-purpose language model through structured, domain-specific instruction. The third is the validation dataset and comparison methodology, which establishes a transparent, reproducible protocol for benchmarking AI-assisted spectral analysis tools against peer-reviewed literature data — a protocol that can be extended and refined in future work.

Scientific integrity requires that the limitations of this work be stated as clearly as its achievements. The tool operates exclusively on transmittance-mode CSV data and does not perform automatic baseline correction, spectral deconvolution, or quantitative analysis. Its peak detection accuracy is highest for sharp, well-resolved bands in organic and hybrid systems, and decreases for broad, overlapping features in purely inorganic spectra. The AI-generated reports, while scientifically coherent and stylistically appropriate, must be treated as informed drafts requiring expert bibliographic completion and critical review before journal submission — they are a starting point, not a final product. The validation exercise, while rigorous within its scope, covers three material systems; a fully comprehensive benchmarking study would require testing across a substantially larger and more diverse spectral library.

The results of this thesis open several directions for future development. In the near term, the integration of automatic asymmetric least-squares baseline correction as an optional preprocessing step would improve peak detection accuracy on spectra with significant background curvature. The addition of a spectral library matching function — comparing detected peaks against a curated database of known materials — would provide a second layer of chemical validation complementary to the AI-generated interpretation. In the medium term, replacing the cloud-based Gemini API with a locally deployed open-source language model would eliminate the dependency on internet connectivity and API quotas, making the tool fully functional in restricted laboratory computing environments. Extension of the system to additional vibrational spectroscopic techniques — Raman spectroscopy in particular, which shares the same peak-based analytical logic — would substantially broaden its user base without requiring fundamental architectural changes.

At a broader level, this work contributes to a growing body of evidence that large language models, when provided with structured numerical input and carefully engineered domain-specific prompts, can generate scientifically reliable and professionally formatted text in specialised technical fields. As these models continue to improve in scientific reasoning and factual grounding, the quality ceiling of AI-assisted spectral interpretation will rise correspondingly, and tools of the kind developed in this thesis will become an increasingly standard component of the research laboratory workflow.

FTIR spectroscopy generates data every day in thousands of laboratories across the world. The gap between that data and the written scientific knowledge it contains has, until now, been bridged almost entirely by human expertise applied manually and repeatedly. This thesis demonstrates that a well-designed intelligent tool can close a meaningful portion of that gap — reliably, transparently, and at no cost to scientific standards. The work is a beginning rather than a conclusion, but it is a beginning grounded in real experimental validation and honest assessment of where the technology stands today and where it can credibly go tomorrow.

References

[1] A. B. D. Nandiyanto, R. Oktiani, and R. Ragadhita, "How to read and interpret FTIR spectroscope of organic material," *Indones. J. Sci. Technol.*, vol. 4, no. 1, pp. 97–118, 2019.

[2] A. Iribarren, E. Hernández-Rodríguez, and L. Maqueira, "Structural, chemical and optical evaluation of Cu-doped ZnO nanoparticles synthesized by an aqueous solution method," *Mater. Res. Bull.*, vol. 60, pp. 376–381, 2014, doi: 10.1016/j.materresbull.2014.08.059.

[3] D. E. Newbury* and N. W. M. Ritchie, "Is scanning electron microscopy/energy dispersive X-ray spectrometry (SEM/EDS) quantitative?," *Scanning*, vol. 35, no. 3, pp. 141–168, 2013.

[4] N. Saadatkhah *et al.*, "Experimental methods in chemical engineering: Thermogravimetric analysis—TGA," *Can. J. Chem. Eng.*, vol. 98, no. 1, pp. 34–43, 2020.

[5] Z. Sabouri *et al.*, "Plant-based synthesis of NiO nanoparticles using salvia macrosiphon Boiss extract and examination of their water treatment," *Rare Met.*, vol. 39, no. 10, pp. 1134–1144, 2020, doi: 10.1007/s12598-019-01333-z.

[6] C. Vogt, C. S. Wondergem, and B. M. Weckhuysen, "Ultraviolet-visible (UV-Vis) spectroscopy," in *Springer handbook of advanced catalyst characterization*, Springer, 2023, pp. 237–264.

[7] M. A. Mohamed, J. Jaafar, A. F. Ismail, M. H. D. Othman, and M. A. Rahman, "Fourier transform infrared (FTIR) spectroscopy," in *Membrane characterization*, Elsevier, 2017, pp. 3–29.

[8] J. Yi *et al.*, "Surface-enhanced Raman spectroscopy: a half-century historical perspective," *Chem. Soc. Rev.*, vol. 54, no. 3, pp. 1453–1551, 2025.

[9] E. L. Schmidt *et al.*, "Near-infrared II fluorescence imaging," *Nat. Rev. Methods Prim.*, vol. 4, no. 1, p. 23, 2024.

[10] X. Li, X. Huang, X. Hu, X. Guo, and Y. Han, "Recent progress on mid-infrared pulsed fiber lasers and the applications," *Opt. Laser Technol.*, vol. 158, p. 108898, 2023.

[11] B. Qin *et al.*, “Far-infrared radiation and its therapeutic parameters: A superior alternative for future regenerative medicine?,” *Pharmacol. Res.*, vol. 208, p. 107349, 2024.

[12] P. Gupta, S. S. Das, and N. B. Singh, “Infrared spectroscopy,” in *Spectroscopy*, Jenny Stanford Publishing, 2023, pp. 151–211.

[13] P. Giacomo, “The michelson interferometer,” *Microchim. Acta*, vol. 93, no. 1, pp. 19–31, 1987.

[14] G. Gomba, A. Parizzi, F. De Zan, M. Eineder, and R. Bamler, “Toward operational compensation of ionospheric effects in SAR interferograms: The split-spectrum method,” *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1446–1461, 2015.

[15] I. M. Mills and A. G. Robiette, “On the relationship of normal modes to local modes in molecular vibrations,” *Mol. Phys.*, vol. 56, no. 4, pp. 743–765, 1985.

[16] S. Kiana, M. Yazdanbakhsh, M. Jamialahmadi, and S. F. Tayyari, “Vibrational assignment and structure of trinuclear oxo-centered of basic formate iron(III) and chromium(III) complexes: A density functional theory study,” *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.*, vol. 130, pp. 287–294, 2014, doi: https://doi.org/10.1016/j.saa.2014.03.130.

[17] B. Mace, “Fundamentals of vibration,” *Fundam. Sound Vib. 2nd ed.; Fahy, F., Thompson, D., Eds*, pp. 311–362, 2015.

[18] H. Yu *et al.*, “Impact of Ultraviolet Radiation on the Aging Properties of SBS-Modified Asphalt Binders,” *Polymers (Basel).*, vol. 11, p. 1111, Jul. 2019, doi: 10.3390/polym11071111.

[19] P. Jayaprakash, M. P. Mohamed, and M. L. Caroline, “Growth, spectral and optical characterization of a novel nonlinear optical organic material: D-Alanine DL-Mandelic acid single crystal,” *J. Mol. Struct.*, vol. 1134, pp. 67–77, Apr. 2017, doi: 10.1016/j.molstruc.2016.12.026.

[20] W. Zhang, C. Wang, and G. S. Kassab, “The mathematical formulation of a generalized Hooke’s law for blood vessels,” *Biomaterials*, vol. 28, no. 24, pp. 3569–3578, 2007.

[21] A. J. Haider, T. Alawsi, M. J. Haider, B. A. Taha, and H. A. Marhoon, “A comprehensive review on pulsed laser deposition technique to effective nanostructure

production: trends and challenges," *Opt. Quantum Electron.*, vol. 54, no. 8, p. 488, 2022, doi: 10.1007/s11082-022-03786-6.

[22] S. R. Rather, M. J. Bezdek, P. J. Chirik, and G. D. Scholes, "Dinitrogen coupling to a terpyridine-molybdenum chromophore is switched on by Fermi resonance," *Chem*, vol. 5, no. 2, pp. 402–416, 2019.

[23] C. J. Chirayil, J. Joy, L. Mathew, M. Mozetic, J. Koetz, and S. Thomas, "Isolation and characterization of cellulose nanofibrils from Helicteres isora plant," *Ind. Crops Prod.*, vol. 59, pp. 27–34, 2014, doi: 10.1016/j.indcrop.2014.04.020.

[24] Q. Carrol, *Radiography in Digital Age*, 3rd ed., vol. 3, no. 1. Springfield, IL, USA: Charles C Thomas Publisher, 2011. [Online]. Available: https://www.ccthomas.com/details.cfm?P_ISBN13=9780398092146

[25] D. F. Swinehart, "The Beer-Lambert law," *J. Chem. Educ.*, vol. 39, no. 7, pp. 333–335, Jul. 1962, doi: 10.1021/ed039p333.

[26] K. H. Vahur Signe, Ivo Leito, "IR Spectroscopy – MOOC: Instrumental analysis of cultural heritage objects," IR Spectroscopy. Accessed: May 30, 2026. [Online]. Available: https://sisu.ut.ee/heritage-analysis/31-ir-spectroscopy/

[27] B. C. Smith, "Infrared Spectroscopy of Polymers, IX: Pendant Ester Polymers and Polycarbonates," *Spectroscopy*, pp. 16-19,31, Dec. 2022, doi: 10.56530/spectroscopy.xn9369p8.

[28] A. Palai, N. R. Panda, and D. Sahu, " Exploring the Photoluminescence Property, Photocatalytic Efficiency, and Antibacterial Activity of Eu-Doped ZnO/SnO 2 Heterostructure ," *ECS J. Solid State Sci. Technol.*, vol. 12, no. 7, p. 076015, 2023, doi: 10.1149/2162-8777/ace84c.

[29] X. Michalet *et al.*, "Quantum dots for live cells, in vivo imaging, and diagnostics," *Science (80-.).*, vol. 307, no. 5709, pp. 538–544, Jan. 2005, doi: 10.1126/science.1104274.

[30] Q. Zhang *et al.*, "Characterization of Chinese vinegars by electronic nose," *Sensors Actuators, B Chem.*, vol. 119, no. 2, pp. 538–546, 2006, doi: 10.1016/j.snb.2006.01.007.

[31] Y. Han, L. Han, Y. Yao, Y. Li, and X. Liu, "Key factors in FTIR spectroscopic analysis of DNA: the sampling technique, pretreatment temperature and sample concentration,"

Anal. Methods, vol. 10, no. 21, pp. 2436–2443, 2018.

[32] T. M. Laid, K. Abdelhamid, L. S. Eddine, and B. Abderrhmane, "Optimizing the biosynthesis parameters of iron oxide nanoparticles using central composite design," *J. Mol. Struct.*, vol. 1229, p. 129497, Apr. 2021, doi: 10.1016/j.molstruc.2020.129497.

[33] J. de O. Primo *et al.*, "Synthesis of Zinc Oxide Nanoparticles by Ecofriendly Routes: Adsorbent for Copper Removal From Wastewater," *Front. Chem.*, vol. 8, p. 571790, 2020, doi: 10.3389/fchem.2020.571790.

[34] S. García-Jimeno and J. Estelrich, "Ferrofluid based on polyethylene glycol-coated iron oxide nanoparticles: Characterization and properties," *Colloids Surfaces A Physicochem. Eng. Asp.*, vol. 420, pp. 74–81, Mar. 2013, doi: 10.1016/j.colsurfa.2012.12.022.

[35] M. Vanaja and G. Annadurai, "Coleus aromaticus leaf extract mediated synthesis of silver nanoparticles and its bactericidal activity," *Appl. Nanosci.*, vol. 3, no. 3, pp. 217–223, Jun. 2013, doi: 10.1007/s13204-012-0121-9.

[36] A. Picard-Lafond and E. Diaz, "Revisiting the liquid mull technique to derive the infrared optical constants of organic powders from transmission IR spectroscopy," in *Chemical, Biological, Radiological, Nuclear, and Explosives (CBRNE) Sensing XXV*, SPIE, 2024, pp. 228–239.

[37] B. C. Smith, *Fundamentals of Fourier transform infrared spectroscopy*. CRC press, 2011.

[38] M. Zoccola, P. Bhavsar, A. Anceschi, and A. Patrucco, "Analytical methods for the identification and quantitative determination of wool and fine animal fibers: a review," *Fibers*, vol. 11, no. 8, p. 67, 2023.

[39] D. Nam, A. Macvean, V. Hellendoorn, B. Vasilescu, and B. Myers, "Using an llm to help with code understanding," in *Proceedings of the IEEE/ACM 46th International Conference on Software Engineering*, 2024, pp. 1–13.

[40] B. Choi, *Introduction to Python Network Automation Volume I-Laying the Groundwork*. Springer, 2024.

[41] C. Miller, *Hands-On Data Analysis with NumPy and pandas: Implement Python packages from data manipulation to processing*. Packt Publishing Ltd, 2018.

[42] A. Sanala, P. Bastian, G. Monkman, V. Kolothodi, and H.-P. Rabl, “A Python-Based Open-Source Gui for Spectral Analysis Using Self-Learning Clustering and Peak Fitting Techniques,” *Available SSRN 5352288*.

[43] R. Gommers *et al.*, “scipy/scipy: SciPy 1.15. 0,” *Zenodo*, 2024.

[44] E. Di Fabrizio *et al.*, “Roadmap on biosensing and photonics with advanced nano-optical methods,” *J. Opt.*, vol. 18, no. 6, p. 63003, 2016.

[45] D. Yadav, *Analytical chemistry foundations*. Educohack Press, 2025.

[46] A. H. Sial, S. Y. S. Rashdi, and A. H. Khan, “Comparative analysis of data visualization libraries Matplotlib and Seaborn in Python,” *Int. J.*, vol. 10, no. 1, pp. 277–281, 2021.

[47] D. Deac, E. Teshome, R. Van Glabbeek, V. Dobrota, A. Braeken, and K. Steenhaut, “Traffic aware scheduler for time-slotted channel-hopping-based IPv6 wireless sensor networks,” *Sensors*, vol. 22, no. 17, p. 6397, 2022.

[48] P. Vijayakumar, M. Pyingkodi, and S. Devi, “Comparative Analysis of AI Chatbot for Assessing Gemini AI, DeepSeek AI, and Qwen AI via OpenRouter API Integration,” in *2025 6th International Conference on Data Intelligence and Cognitive Informatics (ICDICI)*, IEEE, 2025, pp. 1–8.

[49] M. Masalkhi, J. Ong, E. Waisberg, and A. G. Lee, “Google DeepMind’s gemini AI versus ChatGPT: a comparative analysis in ophthalmology,” *Eye*, vol. 38, no. 8, pp. 1412–1417, 2024.

[50] G. Marvin, N. Hellen, D. Jjingo, and J. Nakatumba-Nabende, “Prompt engineering in large language models,” in *International conference on data intelligence and cognitive informatics*, Springer, 2023, pp. 387–402.

[51] G. Nevarez, “Evaluating Large Language Models for Requirements Engineering and Verification: A Comparative Analysis of GPT-5 and Gemini 3 Pro,” 2025, *The University of Texas at El Paso*.

[52] H. K. Lu, “Keeping your API keys in a safe,” in *2014 IEEE 7th International Conference on Cloud Computing*, IEEE, 2014, pp. 962–965.

[53] A. Ohno, “Could Authors of Academic Reports be Discerned Using Formatting Information Obtained by Parsing XML of. docx Documents?,” *IEEJ Trans. Electron.*

Inf. Syst., vol. 143, no. 1, pp. 91–100, 2023.

[54] L. Salah Eddine *et al.*, "Ag/AgCl nanoparticles: photocatalytic degradation of bentazon and adsorption of chromium oxide for sustainable water treatment under optimal pH conditions," *J. Dispers. Sci. Technol.*, pp. 1–13, Sep. 2025, doi: 10.1080/01932691.2025.2556678.

[55] D. Barani *et al.*, "Enhanced antioxidant, anti-inflammatory, and photocatalytic properties of α-tocopherol-coated copper oxide nanoparticles: synthesis, characterisation, and multifunctional applications," *Int. J. Food Sci. Technol.*, vol. 60, no. 1, Jan. 2025, doi: 10.1093/ijfood/vvaf051.

[56] I. Kir, S. E. Laouini, S. Meneceur, A. Bouafia, and H. A. M. Mohammed, "Biosynthesis and characterization of novel nanocomposite ZnO/BaMg2 efficiency for high-speed adsorption of AZO dye," *Biomass Convers. Biorefinery*, vol. 14, no. 16, pp. 19045–19054, 2024, doi: 10.1007/s13399-023-03985-5.

[57] R. Mehrotra, "Infrared Spectroscopy, Gas Chromatography/Infrared in Food Analysis," 2006. doi: 10.1002/9780470027318.a1013.

[58] O. S. Nirmal Ghosh, S. Gayathri, P. Sudhakara, S. K. Misra, and J. Jayaramudu, "Natural Rubber Nanoblends: Preparation, Characterization and Applications BT - Rubber Nano Blends: Preparation, Characterization and Applications," G. Markovic and V. P. M., Eds., Cham: Springer International Publishing, 2017, pp. 15–65. doi: 10.1007/978-3-319-48720-5_2.

APPENDICES

الجمهورية الجزائرية الديمقراطية الشعبية
وزارة التعليم العالي والبحث العلمي
اللجنة الوطنية التنسيقية لمتابعة الإبتكار
وحاضنات الأعمال الجامعية

دليل مشروع

للحصول على شهادة مؤسسة ناشئة
في إطار القرار الوزاري 1275

ديسمبر
2022

CNCSIIU — MHESR — ALGERIA

بطاقة معلومات

حول فريق الاشراف وفريق العمل

1- فريق الاشراف:

فريق الإشراف	
المشرف الرئيسي: د. التجاني محمد العيد	التخصص: هندسة الطرائق

2- فريق العمل:

فريق المشروع	التخصص	الكلية
الطالب: زغيب محمد أصيل	هندسة كيميائية	التكنولوجيا

فهرس المحتويات

فهرس المحتويات

مقدمة

يواجه قطاعا البحث العلمي والمختبرات الصناعية عائقًا اقتصاديًا كبيرًا نتيجة الاعتماد على التحليل اليدوي لأطياف الأشعة تحت الحمراء بتحويل فورييه(FTIR) ، وهي عملية تستغرق وقتًا طويلًا وتكون عرضة للأخطاء البشرية. يقدّم Spectra AIحلاً تجاريًا عالي الجدوى لسد هذه الفجوة السوقية من خلال منصة ذكية سحابية بنظام البرمجيات كخدمة (SaaS) تقوم بأتمتة تفسير البيانات الطيفية وإعداد التقارير العلمية. ومن الناحية الاقتصادية، يساهم هذا الابتكار في خفض التكاليف التشغيلية بشكل كبير، وتحسين تخصيص الموارد ، والاستغناء عن تراخيص البرمجيات التقليدية مرتفعة التكلفة. ويعتمد المشروع على نموذج أعمال قابل للتوسع والاستدامة بدرجة عالية، مع مصادر إيرادات متنوعة تشمل اشتراكات متدرجة موجهة للطلبة والأكاديميين والمختبرات المؤسسية، بالإضافة إلى خيار مرن للدفع مقابل كل عملية تحليل. ومن خلال الاستفادة من الطلب المتزايد بسرعة على تقنيات الذكاء الاصطناعي في مجال الكيمياء التحليلية، يحوّل Spectra AIعملية علمية معقدة إلى مشروع اقتصادي فعّال ومربح يتمتع بإمكانات كبيرة للتوسع والنمو في الأسواق.

المحور الأول

تقديم المشروع

المحور الأول

تقديم المشروع

1. فكرة المشروع (الحل المقترح)

مجال النشاط: التطبيقات الرقمية الحديثة والخدمات العلمية، ويعمل المشروع تحديداً وفق نموذج البرمجيات كخدمة (SaaS) المعتمد على الحوسبة السحابية.

منشأ الفكرة وتطورها: نشأت الفكرة من ملاحظة إحدى العقبات الرئيسية في مجال البحث العلمي، إذ إنه على الرغم من الانتشار الواسع لأجهزة مطيافية الأشعة تحت الحمراء بتحويل فورييه (FTIR)، فإن تفسير الأطياف لا يزال يُنجز يدوياً، وهي عملية تستغرق وقتًا طويلًا، وتكون عرضة للأخطاء، وتتطلب خبرة علمية متخصصة. ومع التطور السريع في تقنيات الذكاء الاصطناعي، تطورت الفكرة نحو استغلال هذه التقنيات لأتمتة هذا المسار التحليلي بالكامل.

ما الذي سيتم إنجازه؟ تطوير منصة ذكية تحمل اسم "Spectra AI"، مصممة لتحليل أطياف FTIR تلقائيًا، والتعرف على المجموعات الوظيفية، وإعداد تقارير علمية جاهزة للنشر بشكل فوري.

كيف سيتم إنجازه؟ سيتم ذلك من خلال دمج خوارزميات متقدمة لمعالجة البيانات والكشف الدقيق عن القمم الطيفية، مع نماذج لغوية كبيرة (LLMs) لتفسير الدلالات الكيميائية للنتائج، وتقديمها عبر واجهة ويب سهلة الاستخدام لا تتطلب تثبيت أي برامج.

من سيقوم بتنفيذه؟ يتم تنفيذ المشروع من طرفي، محمد أصيل زغيب، بالاعتماد على خلفيتي الأكاديمية في هندسة الطرائق ومهاراتي في البرمجة، وذلك تحت إشراف الدكتور تجاني محمد العيد.

أين سيتم تطبيقه؟ يجري حاليًا تطوير المشروع ضمن حاضنة الأعمال بجامعة الشهيد حمه لخضر – الوادي، مع التطلع إلى إطلاقه عبر الإنترنت لخدمة الجامعات ومراكز البحث والمختبرات الصناعية على المستوى العالمي.

2. القيم المقترحة:

- **الابتكار**: (Newness) تقديم حل رقمي مبتكر يعتمد على الذكاء الاصطناعي لأتمتة تفسير أطياف FTIR وإعداد التقارير العلمية بشكل كامل، مما يسد فجوة قائمة في أساليب العمل المخبرية التقليدية التي تعتمد على التحليل اليدوي.
- **الأداء**: (Performance) توفير تحليل سريع ودقيق للغاية للكشف عن القمم الطيفية وتحديد المجموعات الوظيفية، بمستوى أداء وموثوقية يفوقان بشكل ملحوظ أساليب التحليل اليدوي التقليدية.
- **التخصيص**: (Customization) إتاحة خيارات مرنة لتخصيص معايير التحليل، مع إنشاء تقارير علمية تلقائيًا بما يتناسب مع الاحتياجات المختلفة للطلبة والباحثين الأكاديميين والمختبرات الصناعية.
- **إنجاز المهمة**: (Getting the Job Done) تمكين الباحثين والطلبة من إنجاز عملية تحليل البيانات الطيفية المعقدة بسهولة وسرعة، مع الحصول الفوري على نتائج جاهزة للاستخدام أو النشر العلمي.

- **التصميم** :(Design) توفير واجهة ويب حديثة وسهلة الاستخدام وخالية من التعقيد، ومهيأة خصيصاً لعرض البيانات الكيميائية والطيفية بوضوح، بعيداً عن الواجهات التقليدية المعقدة للبرامج القديمة.
- **السعر** :(Price) تقديم نماذج اشتراك مرنة ومتدرجة، تشمل باقات ميسّرة للطلبة والباحثين، وبأسعار أكثر تنافسية مقارنةً بتكاليف تراخيص البرامج التقليدية مرتفعة الثمن.
- **خفض التكاليف** :(Cost Reduction) مساعدة المؤسسات الأكاديمية والمختبرات الصناعية على تقليل التكاليف التشغيلية بشكل كبير من خلال توفير مئات ساعات العمل التي تستهلك عادةً في التفسير اليدوي للبيانات.
- **تقليل المخاطر** :(Risk Reduction) الحد من مخاطر الأخطاء البشرية وسوء تفسير القمم الطيفية والتحيز في النتائج، مما يضمن مخرجات علمية أكثر موثوقية وقابلية للتكرار.
- **إمكانية الوصول** :(Accessibility) توفير المنصة كخدمة سحابية(SaaS) ، مما يتيح الوصول إلى أدوات التحليل الطيفي المتقدمة من أي مكان دون الحاجة إلى أجهزة عالية المواصفات أو عمليات تثبيت معقدة.
- **السهولة وقابلية الاستخدام** :(Convenience / Usability) توفير تجربة استخدام سلسة وبسيطة للغاية، حيث يمكن للمستخدم رفع ملف البيانات الطيفية فقط والحصول فوراً على تحليل شامل وتقرير علمي متكامل.

3. فريق العمل:

أ. فريق المشروع: المهارات والأدوار

يعتمد مشروع Spectra AI على فريق عمل صغير وفعّال مكوّن من شخصين، يجمع بين التنفيذ التقني والإشراف العلمي المتخصص.

• محمد أصيل زغيب (مؤسس المشروع والمطور الرئيسي)

بصفتي طالب ماستر في هندسة الطرائق، أمتلك مهارات تقنية متقدمة في البرمجة (Python وMATLAB)، ومحاكاة العمليات الصناعية باستخدام Aspen HYSYS، بالإضافة إلى توظيف تقنيات الذكاء الاصطناعي وتكاملها في التطبيقات العلمية. كما أنني اجتزت بنجاح برنامج التكوين في ريادة الأعمال بمركز تطوير المقاولاتية (CDE). وتشمل مسؤولياتي التنفيذ الكامل للمشروع، بما في ذلك تطوير البرمجيات، وتصميم واجهة المستخدم، وإعداد نموذج الأعمال، وإدارة العمليات الريادية اليومية.

• الدكتور تجاني محمد العيد (المشرف الأكاديمي والمستشار العلمي)

أستاذ مساعد متخصص في البتروكيمياء، ويتمتع بخبرة واسعة في البحث والتحليل الكيميائي. ويتمثل دوره في الإشراف العلمي على المشروع والتحقق من صحة النتائج العلمية التي تنتجها المنصة، مع ضمان التزام خوارزميات كشف القمم الطيفية والتقارير العلمية المولدة بأعلى المعايير العلمية المعتمدة في المختبرات ومراكز البحث.

ب. التنظيم المناسب: توزيع المهام والمسؤوليات

يعتمد المشروع على هيكل تنظيمي مرن وبسيط يهدف إلى تحقيق أعلى درجات الكفاءة والسرعة في الإنجاز.

• التنفيذ والتطوير التقني

أتولى بشكل كامل جميع أعمال البرمجة، وتطوير الخوارزميات وتدريبها ، وتصميم تجربة وواجهة المستخدم (UI/UX)، إضافة إلى إعداد الاستراتيجية التجارية ونموذج الأعمال. ويسمح هذا النهج المركزي بسرعة تطوير النماذج الأولية، واتخاذ القرارات بمرونة، وتنفيذ التحديثات البرمجية بشكل فوري.

• التحقق العلمي والتوجيه الاستراتيجي

يتولى الدكتور تجاني محمد العيد مسؤولية التقييم العلمي الدقيق لمخرجات الذكاء الاصطناعي، وتوجيه المسار العلمي للمشروع، والتحقق من دقة النتائج والتقارير المولدة، بما يضمن الحفاظ على المصداقية العلمية للمنتج واستخدامه بثقة في الأوساط الأكاديمية والصناعية.

ج. أساليب التفاعل والتواصل

لضمان سير العمل بكفاءة وفعالية، يعتمد الفريق على استراتيجية تواصل هجينة تجمع بين التفاعل المباشر والأدوات الرقمية الحديثة.

• الاجتماعات الحضورية

يتم عقد اجتماعات دورية للتخطيط والتقييم داخل جامعة الشهيد حمه لخضر بالوادي وحاضنة الأعمال التابعة لها ، وذلك لمناقشة مراحل تقدم المشروع، ومراجعة النتائج المحققة، ومعالجة التحديات العلمية والتقنية المعقدة.

• التواصل الرقمي

يتم التواصل المستمر عبر البريد الإلكتروني وتطبيقات المراسلة المختلفة لمراجعة الوثائق الأكاديمية والإدارية، مثل مسودات مذكرة التخرج وملفات ووثائق الحصول على علامة "مشروع مبتكر" وفق القرار الوزاري رقم 1275.

• التعاون السحابي

يتم استخدام منصات التخزين والتطوير السحابية، مثل GitHub ومساحات التخزين المشتركة، لتبادل الشيفرات البرمجية، وقواعد بيانات أطياف FTIR، وملفات المشروع المختلفة بشكل آني ومنظم، مما يسهل التعاون والمتابعة المستمرة لتطور المشروع.

4. أهداف المشروع :

الأهداف التجارية لمشروع Spectra AI

يتمثل الهدف التجاري الشامل لمشروع Spectra AI في ترسيخ مكانة المنصة باعتبارها المساعد الرقمي الرائد والمرجع الأول لتحليل البيانات الطيفية في الأوساط الأكاديمية والصناعية. وباعتماد نموذج البرمجيات كخدمة(SaaS) ، يهدف المشروع إلى استبدال أساليب التحليل اليدوي البطيئة والمعرضة للأخطاء بحل ذكي يعتمد على الذكاء الاصطناعي يتميز بالسرعة والدقة والتكلفة المناسبة.

ولتحقيق نمو مستدام وربحية طويلة الأجل، تم تقسيم الأهداف التجارية والحصص السوقية المستهدفة إلى ثلاث مراحل استراتيجية، استنادًا إلى سوق أولية مستهدفة تضم 13,810 مستخدمًا محتملاً (طلبة، أساتذة، باحثين، ومختبرات).

3.1الأهداف قصيرة المدى (السنة الأولى – دخول السوق والتحقق من جدوى المنتج)

أ. الهدف التجاري

إطلاق المنصة بنجاح، وبناء الوعي بالعلامة التجارية، واستقطاب المستخدمين الأوائل من خلال حملات التسويق الرقمي الموجهة واستراتيجية التجربة المجانية لمدة شهر واحد.

ب. الحصة السوقية المستهدفة

تحقيق معدل اختراق للسوق قدره 10٪، بما يعادل حوالي 1,380 مشتركًا نشطًا من الطلبة والباحثين والمختبرات المتبنية للتقنية في مراحلها الأولى.

ج. التركيز على الإيرادات

توليد تدفقات نقدية أولية بالاعتماد أساسًا على اشتراكات الطلبة منخفضة التكلفة وباقات الباحثين الأفراد، بهدف التحقق من توافق المنتج مع احتياجات السوق.(Product-Market Fit)

3.2 ##الأهداف متوسطة المدى (من السنة الثانية إلى السنة الثالثة – التوسع والاعتماد المؤسسي)

د. الهدف التجاري

الانتقال من نموذج الاشتراكات الفردية (B2C) إلى تأمين عقود مؤسسية عالية القيمة(B2B) ، من خلال دمج المنصة في البرامج التعليمية الجامعية وإبرام شراكات رسمية مع مراكز البحث العلمي.

هـ. الحصة السوقية المستهدفة

توسيع قاعدة المستخدمين بشكل متسارع للوصول إلى نسبة تتراوح بين 30٪ و60٪ من السوق المستهدفة، أي ما يزيد عن 8,200 مستخدم نشط بحلول نهاية السنة الثالثة.

و. التركيز على الإيرادات

تعظيم الإيرادات من خلال تراخيص المؤسسات والمختبرات (باقة المخابر) مع تعزيز خدمة الدفع مقابل التحليل (Pay-per-Analysis) للمستخدمين غير المنتظمين.

3.3 ##الأهداف طويلة المدى (من السنة الرابعة إلى الخامسة وما بعدها – الريادة والتوسع الدولي)

ز. الهدف التجاري

تحقيق الريادة في السوق المحلية والانطلاق نحو التوسع الدولي، خاصة في منطقة الشرق الأوسط وشمال إفريقيا (MENA) والمؤسسات الأكاديمية العالمية. كما يسعى المشروع إلى توسيع قدرات المنصة لتشمل تقنيات تحليل طيفي أخرى إلى جانبFTIR ، مثل:

التحليل الطيفي بالأشعة فوق البنفسجية والمرئية.(UV-Vis)

حيود الأشعة السينية.(XRD)

تقنيات التحليل الطيفي المتقدمة الأخرى.

ح. الحصة السوقية المستهدفة

الوصول إلى معدل تبنٍ يتراوح بين 80٪ و100٪ داخل السوق المحلية الأساسية، مع تحقيق حضور ملموس وقابل للقياس في سوق البرمجيات العلمية الدولية.

ط. التركيز على الإيرادات

بناء مصدر إيرادات متكرر ومستقر وقابل للتنبؤ به، مع هوامش ربح مرتفعة، بالاعتماد على عمليات الذكاء الاصطناعي المؤتمتة وانخفاض التكاليف الحدية المرتبطة بإضافة مستخدمين جدد إلى المنصة.

5. جدول زمني لتحقيق المشروع :

الشهر أو الأسبوع

			1	2	3	4	5	6	7
1		جمع البيانات وتصميم البنية المعمارية :جمع ملفات أطياف FTIR الدقيقة والموثوقة لضمان الحصول على نتائج تحليل عالية الجودة، مع تصميم البنية المعمارية العامة للنظام بما يضمن الكفاءة وقابلية التوسع.	✓	✓					
2		خط أنابيب إدخال ومعالجة البيانات: تطوير بيئة المعالجة الرقمية الخاصة بالمنصة، وإعداد خط أنابيب متكامل لاستقبال البيانات الطيفية ومعالجتها مسبقًا وتنظيفها قبل بدء التحليل.		✓	✓				
3		الكشف الآلي عن القمم الطيفية: توظيف تقنيات الذكاء الاصطناعي لاستخراج القمم الطيفية تلقائيًا وتحليل المركبات الكيميائية المرتبطة بها ، من خلال تطوير خوارزمية متقدمة للكشف الآلي عن القمم بدقة عالية.		✓	✓	✓			
4		وحدة إنشاء التقارير العلمية: إعداد تقارير علمية آلية تتوافق مع احتياجات مختلف فئات المستخدمين، وذلك عبر تطوير وحدة ذكية تعتمد على الذكاء الاصطناعي لإنشاء التقارير والوثائق العلمية بصورة تلقائية ومنظمة.			✓	✓	✓		
5		التحقق التجريبي: تقييم والتحقق من أداء الأداة من خلال مقارنة نتائجها ببيانات أطياف FTIR المنشورة والمعتمدة في الأدبيات العلمية، وذلك لضمان دقة التحليل وموثوقية النتائج.						✓	
66		المخرجات النهائية والنشر: عرض نتائج التحليل والتقارير العلمية بطريقة واضحة ومنظمة وسهلة الفهم ، مع توفيرها بصيغ قابلة للتنزيل أو الطباعة، بما يتيح للمستخدمين الاستفادة منها مباشرة في الأبحاث والتقارير الأكاديمية والمهنية.							✓

المحور الثاني

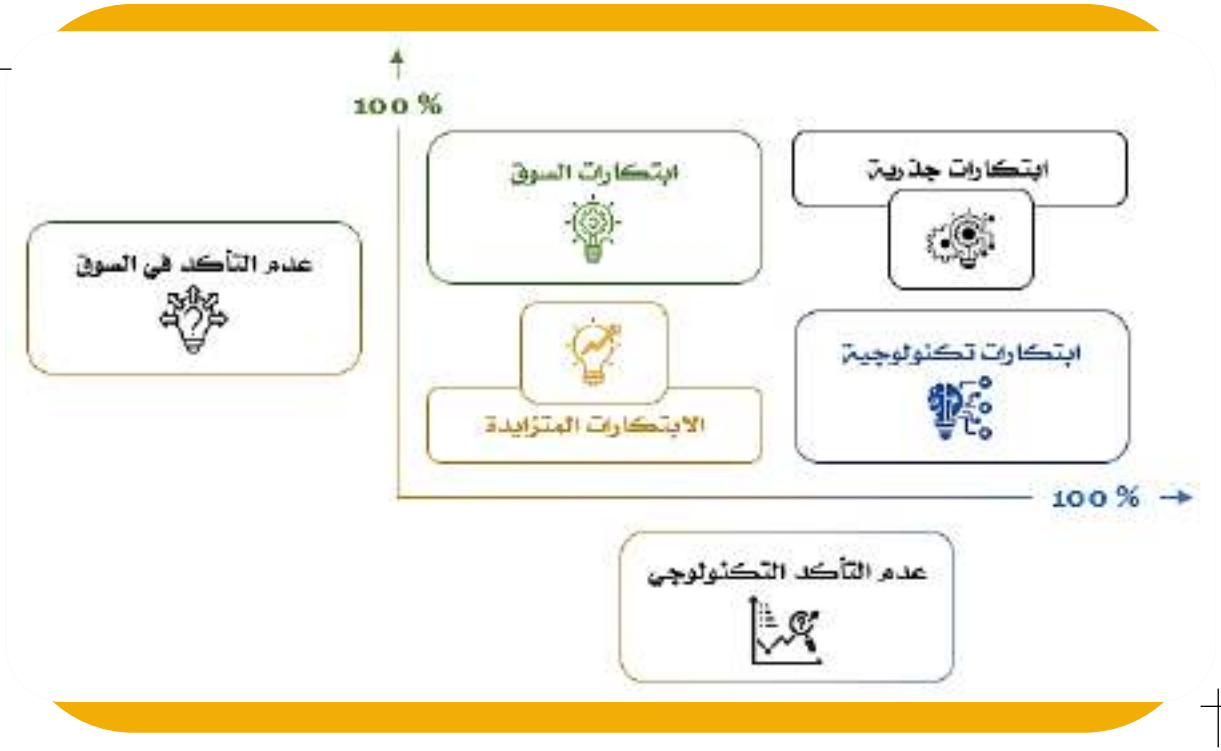
الجوانب الإبتكارية

المحور الثاني

الجوانب الابتكارية

1. طبيعة الابتكارات :

ينبغي أن يحدد هنا طبيعة الابتكارات المعتمدة في المشروع :



2. مجالات الابتكارات :

- **يمكن أن يتجسد الابتكار في مشروع Spectra AI من خلال المجالات التالية:**
- **عمليات جديدة (New Processes): زيادة الكفاءة التشغيلية ورفع إنتاجية المختبرات من خلال أتمتة المهام التقليدية التي تنجز يدوياً، مثل تفسير أطياف FTIR وإعداد التقارير العلمية، وهي عمليات تستغرق وقتًا طويلًا وتتطلب خبرة متخصصة.**
- **تجارب جديدة (New Experiences): تعزيز القيمة المقدمة للمستخدمين في الأوساط الأكاديمية والصناعية عبر تحويل عملية تحليل البيانات الطيفية الخام إلى تجربة رقمية تفاعلية وسلسة، مدعومة برؤى وتفسيرات فورية يولدها الذكاء الاصطناعي.**
- **خصائص جديدة (New Features): تقديم خدمات رقمية متطورة من خلال دمج مزايا متقدمة مثل الكشف الآلي عن القمم الطيفية، وإمكانية ضبط حساسية التحليل، وإنشاء مستندات وتقارير علمية بصيغة Word جاهزة للاستخدام في الأبحاث والمنشورات العلمية.**
- **عملاء جدد (New Customers): توسيع نطاق الاستفادة من أدوات التحليل المتقدمة لتشمل الباحثين غير المتخصصين، وطلبة الدراسات العليا في المراحل الأولى، وموظفي مراقبة**

الجودة، الذين يمكنهم الاستفادة من التحليل المبسط والموجّه دون الحاجة إلى خبرة عميقة في التحليل الطيفي.

- عروض جديدة (New Offerings): تقديم منتج مبتكر وفريد من نوعه في السوق يتمثل في مساعد ذكي يربط بين المخرجات الخام لأجهزة التحليل الطيفي والنصوص العلمية المنظمة، مما يختصر مراحل متعددة من العمل البحثي في منصة واحدة.
- نماذج أعمال جديدة (New Models): إحداث تحول في أسلوب تقديم القيمة للمستخدمين من خلال الانتقال من البرمجيات التقليدية المكلفة والمقيدة بالأجهزة المحلية إلى نموذج البرمجيات كخدمة (SaaS) القائم على الاشتراكات المرنة وخيارات الدفع مقابل كل عملية تحليل، مما يجعل الخدمة أكثر سهولة في الوصول وأكثر ملاءمة لمختلف فئات المستخدمين.).

المحور الثالث

التحليل الاستراتيجي للسوق

المحور الثالث

التحليل الاستراتيجي للسوق

1. عرض القطاع السوقي:

السوق المستهدفة وإمكانات التعاقد

• السوق المحتملة (Potential Market)

تتكون السوق المحتملة من الأفراد والمؤسسات التي تعتمد بشكل كبير على التحليل الطيفي في أنشطتها العلمية والبحثية والصناعية. وتشمل هذه السوق طلبة الدراسات العليا، والباحثين الأكاديميين، والمختبرات الصناعية (في مجالات الكيمياء، والمواد، والصيدلة، والطاقة)، بالإضافة إلى المختبرات الطبية ومراكز البحث العلمي.

• السوق المستهدفة (Target Market Segments)

1. طلبة الدراسات العليا (الماستر والدكتوراه):

تمثل هذه الفئة أكبر شريحة مستهدفة، حيث يُقدّر عدد المستخدمين المحتملين بحوالي 11,000 مستخدم. وتتمثل احتياجاتهم في الحصول على أدوات ذكية توفر الوقت والجهد في تحليل البيانات الطيفية وإعداد التقارير العلمية.

2. الباحثون الأكاديميون والأساتذة الجامعيون:

يُقدّر عددهم بحوالي 2,700 مستخدم في الجامعات ومخابر البحث. ويحتاجون إلى أدوات موثوقة تقلل من الأخطاء البشرية وتسرّع عملية تفسير النتائج واستخلاص الاستنتاجات العلمية.

3. المختبرات الصناعية:

تمثل قطاعًا مهمًا يركز على تحسين الكفاءة التشغيلية وتسريع عمليات مراقبة الجودة وتحليل المواد في مختلف الصناعات الكيميائية والدوائية والطاقوية وغيرها.

4. المختبرات الطبية ومراكز البحث:

يُقدّر حجم هذه الفئة الأولية بحوالي 110 مختبرات بحثية تحتاج إلى تحليلات طيفية دقيقة ومتقدمة لدعم أنشطتها البحثية والتطبيقية.

• أسباب استهداف هذه السوق

تم اختيار هذه السوق لوجود حاجة واضحة إلى حل ذكي وسهل الاستخدام وبتكلفة مناسبة، قادر على أتمتة تحليل أطياف FTIR وإنتاج تقارير علمية دقيقة بسرعة وكفاءة. فمعظم الأساليب الحالية لا تزال تعتمد على التفسير اليدوي الذي يتطلب وقتًا طويلًا وخبرة متخصصة، في حين أن البرامج المتاحة غالبًا ما تكون معقدة ومكلفة ولا توفر إمكانيات متقدمة لإنشاء التقارير العلمية تلقائيًا.

• إمكانية الحصول على العقود

تتمتع المنصة بإمكانية كبيرة للحصول على عقود وشراكات طويلة الأمد، خاصة في القطاع الأكاديمي الذي يضم الطلبة والباحثين، من خلال نموذج البرمجيات كخدمة (SaaS) القائم على الاشتراكات المتدرجة، مثل:

باقة الطلبة (Student Package)

باقة الباحثين (Researcher Package)

باقة المؤسسات والمختبرات (Enterprise Package)

كما تتيح باقة المؤسسات والمختبرات إمكانية منح تراخيص استخدام شاملة للمختبرات والجامعات ومراكز البحث، مما يعزز فرص إبرام عقود مؤسسية مستدامة ويوفر مصدر دخل متكرر ومستقر للمشروع على المدى الطويل.

2. قياس شدة المنافسة:

أ. المنافسون المباشرون وغير المباشرين

•المنافسون المباشرون (برامج التحليل الطيفي التجارية)

- تشمل هذه الفئة حزم البرمجيات التجارية التقليدية الخاصة بالتحليل الطيفي، مثل

: PerkinElmer Spectrum و Bruker OPUS

وتعد هذه البرامج الأدوات الافتراضية المستخدمة في المختبرات، حيث تكون غالبًا مدمجة مع أجهزة التحليل الطيفي وتستخدم بشكل واسع في البيئات الأكاديمية والصناعية.

•المنافسون غير المباشرين (النماذج اللغوية العامة)

تشمل النماذج اللغوية الكبيرة متعددة الاستخدامات، مثل: ChatGPT و Claude

حيث يلجأ بعض الباحثين إلى استخدامها بصورة غير مباشرة لتحليل أو تفسير البيانات الطيفية بعد تحويلها إلى نصوص أو جداول، رغم أنها ليست مصممة خصيصًا لهذا الغرض.

•المنافسون غير المباشرين (التفسير اليدوي التقليدي)

يمثل التفسير اليدوي للأطياف الأسلوب التقليدي السائد ، حيث يعتمد الباحثون على مطابقة القمم الطيفية مع الجداول المرجعية المنشورة والأدبيات العلمية المتخصصة لاستخلاص النتائج وتحليل المركبات.

ب. التواجد والحصة السوقية

•البرامج التجارية (PerkinElmer Spectrum وBruker OPUS)

تتمتع هذه البرمجيات بحصة سوقية مهيمنة وشبه احتكارية داخل الجامعات والمختبرات الصناعية. ويرجع ذلك إلى ارتباطها المباشر بأجهزة التحليل الطيفي التي تباع معها، مما يجعلها موجودة في نسبة كبيرة جداً من مختبرات توصيف المواد والتحليل الكيميائي حول العالم.

•النماذج اللغوية العامة وأساليب العمل اليدوية

على الرغم من أن النماذج اللغوية العامة لا تمتلك حصة سوقية رسمية في مجال برمجيات التحليل الطيفي المتخصصة، فإن استخدامها يتزايد تدريجيًا في المراحل اللاحقة لمعالجة البيانات وكتابة التقارير العلمية. أما التفسير اليدوي المعتمد على المراجع العلمية فلا يزال يشكل الأسلوب الأكثر انتشارًا لدى العديد من المجموعات البحثية الصغيرة والمتوسطة في مختلف أنحاء العالم، خاصة في المؤسسات التي لا تمتلك حلولًا برمجية متقدمة أو مخصصة للتحليل الطيفي.

3. الاستراتيجيات التسويقية

يعتمد مشروع Spectra AI على استراتيجية تسويق رقمية منخفضة التكلفة وعالية التأثير، تهدف إلى جذب المستخدمين وبناء قاعدة عملاء مستدامة بأقل قدر ممكن من النفقات التسويقية.

1.النهج الاستراتيجي (فعّال من حيث التكلفة)

•**التواصل العضوي**(Organic Outreach) @

الاستفادة من المنصات الأكاديمية والمهنية مثل ResearchGate**و** LinkedIn**للتواصل المباشر مع الباحثين والطلبة والمختصين في المجالات العلمية، مما يسمح بالوصول إلى الفئة المستهدفة بدقة عالية.**

•**النمو القائم على المحتوى**(Content-Led Growth)

توفير محتوى تعليمي وشروحات تقنية ودروس تطبيقية تساعد المستخدمين على فهم التحليل الطيفي والاستفادة من إمكانات المنصة، مما يعزز الثقة بالمنتج ويُظهر القيمة العملية للتحليل المؤتمت المدعوم بالذكاء الاصطناعي.

•**الشراكات المؤسسية**(Institutional Alliances)

إقامة شراكات مع حاضنات الأعمال الجامعية ومراكز دعم الابتكار ونقل التكنولوجيا(TISC) **، بهدف تسريع انتشار المنصة محليًا وتعزيز تبنيها داخل المؤسسات الأكاديمية والبحثية.**

.2**المزيج التسويقي المتوازن**(4Ps)

•**المنتج**(Product)

منصة سحابية ذكية توفر خدمات الكشف الآلي عن القمم الطيفية وتحليل أطيافFTIR **، مع إنشاء تقارير علمية تلقائية مدعومة بالذكاء الاصطناعي.**

•**السعر**(Price)

اعتماد نموذج البرمجيات كخدمة (SaaS) **من خلال اشتراكات متدرجة تناسب مختلف فئات المستخدمين:**

- **باقة الطلبة.**
- **باقة الباحثين.**
- **باقة المؤسسات والمختبرات.**

بالإضافة إلى خيار الدفع مقابل كل عملية تحليل (Pay-per-Analysis) **لضمان المرونة وسهولة الوصول إلى الخدمة.**

•**المكان**(Place)

تقدَّم الخدمة عبر منصة ويب سحابية لا تتطلب أي تثبيت أو تجهيزات خاصة، مما يتيح الوصول إليها عالميًا وبشكل فوري من أي جهاز متصل بالإنترنت.

•**الترويج**(Promotion)

الاعتماد على وسائل ترويج متخصصة وموجهة تشمل:

- **الندوات التقنية عبر الإنترنت.**(Webinars)
- **ورش العمل العلمية والمؤتمرات الأكاديمية.**
- **حملات التواصل الرقمي المباشر مع الباحثين والمختبرات والمؤسسات العلمية.**

وتهدف هذه الأنشطة إلى التعريف بالمنصة، وإبراز مزاياها التقنية، وتشجيع تبنيها من قبل الفئات المستهدفة.

المحور الرابع

خطة الإنتاج والتنظيم

المحور الرابع

خطة الإنتاج والتنظيم

1. عملية الإنتاج :

أ. جمع البيانات (المواد الأولية)

- **الوصف :** جمع بيانات أطياف الأشعة تحت الحمراء بتحويل فورييه (FTIR) عالية الدقة والجودة بصيغ رقمية مثل CSV و TXT.
- **ضمان الجودة :** التحقق من استقرار خط الأساس (Baseline) وسلامة الإشارة الطيفية وجودتها قبل بدء عملية التحليل، لضمان الحصول على نتائج دقيقة وموثوقة.

ب. المعالجة والتفسير (عملية التصنيع)

- **الوصف :** استخدام خوارزميات الذكاء الاصطناعي للكشف التلقائي عن القمم الطيفية، وتحديد المجموعات الوظيفية، وتفسير البنى الجزيئية للمركبات المدروسة.
- **ضمان الجودة :** تقليل الأخطاء البشرية من خلال اعتماد منهجية تحليل موحدة قائمة على الخوارزميات الذكية، مما يضمن الاتساق والدقة في النتائج.

ج. تخصيص النتائج (مواءمة المنتج مع احتياجات المستخدم)

- **الوصف :** تنظيم البيانات التحليلية الخام وتحويلها إلى مخرجات تتوافق مع المتطلبات البحثية الخاصة بكل مستخدم أو مشروع علمي.
- **ضمان الجودة :** تعزيز دقة النتائج وملاءمتها من خلال تكييف التفسير العلمي وفق طبيعة المادة أو العينة محل الدراسة.

د. إنشاء التقارير وتصديرها (التغليف)

- **الوصف :** إنشاء تقرير علمي منظم بصورة تلقائية يتوافق مع المعايير الأكاديمية ومتطلبات المجلات العلمية، ويتضمن قسم "النتائج والمناقشة (Results and Discussion)" بشكل جاهز للاستخدام أو النشر.
- **ضمان الجودة :** التأكد من سلامة التنسيق، ودقة المحتوى العلمي، واتساق التقرير مع البيانات التحليلية المستخرجة، بما يضمن إنتاج وثائق احترافية عالية الجودة قابلة للتنزيل والطباعة.

يمكن الاستعانة بمخطط يشرح مراحل عملية الإنتاج :

2. النموذج:

أ. سياسة المشتريات (الموارد والأصول)

تنقسم سياسة المشتريات في المشروع إلى ثلاث فئات استراتيجية رئيسية:

•المواد الأولية (البيانات والمكتبات الطيفية)

نعطي الأولوية للحصول على قواعد بيانات وأطياف FTIR عالية الجودة وموثقة علمياً من مصادر محكمة، بهدف تدريب نماذج الذكاء الاصطناعي وتحسين أدائها باستمرار. وتعتمد عملية الاختيار على سلامة البيانات، والتوافق مع سياسات الوصول المفتوح، ومدى ارتباطها بفئات المواد المستهدفة.

•البنية التحتية واللوازم (الحوسبة السحابية)

يتم اقتناء خدمات الحوسبة السحابية والموارد الحاسوبية القابلة للتوسع لاستضافة منصة Spectra AI وتشغيلها. وتركز سياسة الاختيار على ضمان التوافر العالي للخدمة، والامتثال لمعايير الأمن السيبراني، وإمكانيات التوسع التلقائي لمواكبة تغيرات الطلب وعدد المستخدمين.

•التجهيزات (أصول التطوير)

يتم إعطاء الأولوية للأجهزة والمعدات عالية الأداء المستخدمة في تطوير البرمجيات، وإدارة مستودعات الشيفرات البرمجية، وتشغيل بيئات التكامل والتسليم المستمر (CI/CD).

ب. الموردون الرئيسيون

يتم تصنيف الموردين وفقًا لدورهم في سلسلة القيمة الأساسية للمشروع:

•مزودو الخدمات السحابية (Cloud Service Providers - CSPs)

مثل AWS و Google Cloud و Microsoft Azure، حيث يمثلون شركاء أساسيين لاستضافة المنصة، وإدارة قواعد البيانات، ونشر نماذج الذكاء الاصطناعي.

•

الشركاء الأكاديميون ومراكز البحث

تشمل الجامعات ومراكز البحث العلمي التي توفر مكتبات بيانات طبيعية موثقة ومعتمدة تستخدم في معايرة النماذج وتحسين دقتها.

•مزودو واجهات البرمجة(API Providers)

وهم الموردون الذين يوفرون واجهات النماذج اللغوية الكبيرة، مثل Google AI Studio وGemini، والتي تستخدم لتشغيل وحدة إنشاء التقارير العلمية المعتمدة على الذكاء الاصطناعي.

ج. سياسة الدفع وشروط الاستلام

•سياسة الدفع

يعتمد المشروع نموذج الدفع القائم على الاشتراك بالنسبة للخدمات السحابية وواجهات البرمجة، باعتبارها مصاريف تشغيلية(OPEX) ، وذلك لمواءمة التكاليف مع نمو الإيرادات. كما يتم تفضيل دورات الفوترة الشهرية أو السنوية للاستفادة من الخصومات وتحسين التدفقات النقدية.

•شروط التسليم والاستلام

- **الخدمات السحابية وواجهات البرمجة**:يتم توفير الخدمة بشكل فوري ومستمر بمجرد إتمام عملية الدفع وتفعيل مفاتيح الوصول(API Keys).
- **مكتبات البيانات**:يتم تسليمها عبر قنوات نقل رقمية آمنة، مثل مستودعات S3 أو قواعد البيانات المحمية، بعد إبرام اتفاقيات الشراكة.
- **أصول التطوير**:تتم عمليات الشراء وفق إجراءات أوامر الشراء(Purchase Orders) ، مع فترات سداد اعتيادية تصل إلى 30 يومًا بعد استلام المعدات أو البرمجيات والتأكد من سلامتها.

3. اليد العاملة :

•إمكانات خلق مناصب العمل

من المتوقع أن يساهم المشروع في استحداث فرص عمل متخصصة مع توسع النشاط، ومن أبرزها:

الفريق التقني الأساسي

- مهندسو الذكاء الاصطناعي.
- مطورو البرمجيات وتطبيقات الويب.
- مختصو صيانة وتطوير الخوارزميات.

الخبراء المتخصصون

- كيميائيون تحليليون لضمان جودة النتائج العلمية ودقة التفسيرات الطيفية.

التشغيل والدعم

- مختصون في التسويق الرقمي.
- مسؤولو اكتساب العملاء.
- فرق الدعم التقني وخدمة المستخدمين.

•ملفات الموارد البشرية ومكان العمل

يتطلب المشروع فريقًا عالي الكفاءة يجمع بين الخبرة التقنية والعلمية.

الملفات المطلوبة

- مطورو البرمجيات (الويب والذكاء الاصطناعي).
- علماء البيانات.
- الكيميائيون التحليليون.

مكان العمل

على الرغم من أن مراحل التطوير الأولى تتم داخل البيئة الجامعية، فإن طبيعة المنصة باعتبارها خدمة سحابية (SaaS) تسمح باعتماد نموذج عمل موزع يعتمد على البنية التحتية السحابية وأدوات التعاون عن بُعد.

a. **إمكانيات الاستعانة بمصادر خارجية**(Outsourcing)

يعتمد المشروع نموذج تشغيل مرنًا يسمح بالاستعانة بمصادر خارجية في بعض الأنشطة غير الأساسية.

البنية التحتية السحابية

إسناد خدمات الاستضافة والحوسبة السحابية إلى مزودين متخصصين.

خدمات الدعم

إمكانية التعاقد مع جهات خارجية لتنفيذ حملات التسويق المتقدمة أو إدارة بعض قنوات الدعم الفني الثانوية.

التطوير الأساسي

الاحتفاظ داخليًا بالسيطرة الكاملة على الخوارزميات الذكية والمنطق العلمي الخاص بتحليل الأطياف، باعتبارها تمثل الملكية الفكرية الجوهرية للمشروع.

4. الشراكات الرئيسية :

- **المؤسسات الأكاديمية ومراكز البحث**

تمثل الجامعات ومراكز البحث الحاضنة الأساسية للابتكار، حيث توفر البيئة العلمية اللازمة لتطوير المنصة واختبارها ميدانيًا في مراحلها الأولى.

- **الشبكات العلمية والمهنية**

يُعد التعاون مع الباحثين والمتخصصين في الكيمياء التحليلية، إضافة إلى الاستفادة من المنصات المهنية والعلمية مثل LinkedIn وResearchGate، عنصرًا أساسيًا للتحقق من فعالية الأداة، وتعزيز المصداقية العلمية، والوصول إلى الفئات المستهدفة من المستخدمين.

- **مزودو التكنولوجيا والبنية التحتية**

يشكل مزودو خدمات الحوسبة السحابية شركاء استراتيجيين لاستضافة منصة البرمجيات كخدمة (SaaS)وضمان قابليتها للتوسع والاعتمادية العالية.

- **التوجيه والدعم المتخصص**

يستفيد المشروع من خدمات مركز دعم التكنولوجيا والابتكار (TISC) وحاضنات الأعمال الجامعية، التي توفر التوجيه الاستراتيجي والدعم المتخصص في مجالات الابتكار وريادة الأعمال وتطوير المشاريع الناشئة.

المحور الخامس

الخطة المالية

المحور الخامس

الخطة المالية

1. التكاليف والأعباء :

1. التكاليف والمصاريف

1. الاحتياجات المالية وميزانية الاستثمار

يتطلب المشروع ميزانية منظمة ومدروسة للانتقال من نموذج أكاديمي أولي إلى منتج تجاري متكامل يعمل وفق نموذج البرمجيات كخدمة (SaaS). وتشمل أهم عناصر التكلفة ما يلي:

•البحث والتطوير

التكاليف المرتبطة بتطوير وتحسين خوارزميات الذكاء الاصطناعي، ورفع كفاءة الكشف عن القمم الطيفية، وتطوير البنية البرمجية السحابية للمنصة.

•البنية التحتية والتشغيل

تشمل نفقات الاستضافة السحابية) مثل AWS أو Azure أو(Google Cloud ، ورسوم استخدام واجهات البرمجة والتكامل مع نماذج الذكاء الاصطناعي) مثل(Gemini API ، بالإضافة إلى تكاليف تشغيل وصيانة منصة آمنة وقابلة للتوسع.

•التسويق واكتساب العملاء

تخصيص ميزانية للحملات التسويقية الرقمية عبر منصات مثل LinkedIn و ResearchGate والندوات الإلكترونية (Webinars)، بهدف زيادة الوعي بالمنصة وجذب المستخدمين الجدد.

•المصاريف الإدارية والقانونية

تشمل تكاليف التراخيص، والامتثال للمتطلبات القانونية والتنظيمية، بالإضافة إلى المصاريف التشغيلية والإدارية المستمرة.

2. مصادر التمويل واستراتيجية التمويل

يعتمد المشروع على مزيج تمويلي مستدام يهدف إلى دعم النمو الأولي وضمان الاستمرارية على المدى الطويل.

•دعم حاضنة الأعمال الجامعية

الاستفادة من البنية التحتية والدعم المؤسسي الذي توفره جامعة الشهيد حمه لخضر ومركز تطوير المقاولاتية (CDE)، سواء من خلال الموارد أو التوجيه والإرشاد.

•التمويل الذاتي

الاعتماد على الموارد الذاتية لفريق المشروع خلال المراحل الأولى من التطوير والاختبار.

•المنح وبرامج دعم الابتكار

الاستفادة من برامج الدعم الحكومية والوزارية المخصصة لإنشاء المؤسسات المصغرة والمشاريع المبتكرة والشركات الناشئة.

•الشراكات الاستراتيجية

التعاون مع المؤسسات الأكاديمية والصناعية لتطوير بعض الخصائص والخدمات مقابل منحهم وصولاً مبكراً للمنصة أو تراخيص استخدام خاصة.

3.نموذج الإيرادات واسترداد الاستثمار (العائد على الاستثمار)

يعتمد المشروع على نموذج إيرادات مستدام ومتدرج يهدف إلى تحقيق تدفقات نقدية منتظمة وقابلة للنمو.

•الإيرادات القائمة على الاشتراكات (شهريًا أوسنويًا)

باقة الطلبة(Student Tier)

اشتراك منخفض التكلفة يتيح للطلبة الاستفادة من خدمات المنصة الأساسية.

باقة الباحثين(Researcher Tier)

اشتراكات شهرية أوسنوية موجهة للباحثين والأكاديميين الأفراد.

باقة المؤسسات والمختبرات(Enterprise/Laboratory Tier)

تراخيص شاملة موجهة للمختبرات الجامعية ومراكز البحث والتطوير والمؤسسات الصناعية.

•الإيرادات القائمة على الاستخدام

توفير خدمة "الدفع مقابل التحليل(Pay-per-Analysis) "، والموجهة للمستخدمين العرضيين أو الجهات التي تحتاج إلى عدد محدود من التحاليل دون اشتراك دائم.

•الخدمات ذات القيمة المضافة

تحقيق إيرادات إضافية من خلال تقديم خدمات متقدمة ومخصصة لإعداد التقارير العلمية والتحليلات المتخصصة للمشاريع الصناعية المعقدة.

2. رأس المال :

أ. السيناريو المتفائل (معدل تبنٍ مرتفع)

يفترض هذا السيناريو تحقيق انتشار سريع في السوق، وإبرام شراكات ناجحة مع الجامعات ومراكز البحث الكبرى، إلى جانب المحافظة على معدل مرتفع من العملاء المؤسسيين.

العوامل الدافعة

ارتفاع معدل تحويل المستخدمين من النسخة التجريبية المجانية إلى الاشتراكات المدفوعة.

نجاح الحملات التسويقية عبر LinkedIn و ResearchGate.

التبني السريع للمنصة من قبل المختبرات الصناعية ومراكز البحث.

التأثير المتوقع

الوصول إلى الإيرادات الشهرية المتوقعة أو تجاوزها، والمقدرة بحوالي 33,790,000 دينار جزائري شهريًا بحلول السنة الثالثة، مدعومة بتحقيق هدف تبنٍ سوقي يصل إلى 60٪ من السوق المستهدفة.

ب. السيناريو المتحفظ (اختراق سوقي أبطأ)

يفترض هذا السيناريو مواجهة ضغوط تنافسية من البرمجيات التحليلية التقليدية، بالإضافة إلى بطء عمليات التبني المؤسسي واتخاذ القرار داخل المؤسسات الأكاديمية والصناعية.

العوامل المؤثرة

طول دورة اتخاذ القرار والتعاقد داخل الجامعات والمؤسسات البحثية.

القيود المالية والميزانيات المحدودة لدى المختبرات الصغيرة.

المنافسة السعرية القوية من البرمجيات الراسخة في السوق.

التأثير المتوقع

تحقيق نمو تدريجي وأكثر تحفظًا، بحيث تبقى معدلات التبني أقل من 20٪ إلى 30٪ خلال السنوات الثلاث الأولى، مما قد يستدعي إعادة تقييم سياسة التسعير، وتعزيز الاعتماد على نموذج الدفع مقابل التحليل لضمان استقرار التدفقات النقدية واستدامة النشاط.

DETAIL CHIFFRE D'AFFAIRE **STARTUP : Spectra AI**

	REALIZED			FORECAST		
User Segments	N-2	N-1	N	N+1	N+2	N+3
Market Adoption Rate	-	-	-	10%	30%	60%
Active Student Subscribers	-	-	-	1,100	3,300	6,600
Active Researcher Subscribers	-	-	-	270	810	1,620
Active Enterprise/Lab Accounts	-	-	-	11	33	66
Average Revenue per User (Monthly DZD)	-	-	-	2,000	4,000	9,000
Total Monthly Revenue (DZD)	-	-	-	3,379,000	10,137,000	20,274,000

3. جدول حسابات النتائج المتوقع :

أ. قائمة الدخل (الأرباح والخسائر)

تُعد قائمة الدخل أداة مالية أساسية تُستخدم لتقديم ملخص سنوي لأداء مشروع Spectra AI من خلال مقارنة الإيرادات المحققة بالمصاريف التشغيلية المتكبدة خلال الفترة المالية.

وتشمل الإيرادات الرئيسية:

عائدات اشتراكات الطلبة.

عائدات اشتراكات الباحثين والأكاديميين.

إيرادات تراخيص المؤسسات والمختبرات.

إيرادات خدمة الدفع مقابل التحليل.

إيرادات الخدمات الاستشارية والتقارير المتخصصة.

أما المصاريف التشغيلية فتشمل:

تكاليف استخدام واجهات الذكاء الاصطناعي (API).

تكاليف الاستضافة والحوسبة السحابية.

مصاريف البحث والتطوير.

نفقات التسويق واكتساب العملاء.

المصاريف الإدارية والتشغيلية.

وتُظهر قائمة الدخل بوضوح صافي الربح أو الخسارة خلال كل فترة مالية، مما يسمح بتقييم الأداء الاقتصادي للمشروع وقياس مدى نجاحه في الانتقال من مرحلة الاستثمار والتطوير الأولية إلى مرحلة تحقيق الربحية المستدامة والنمو التجاري طويل الأجل.

ب. احتياج رأس المال العامل (WCR / BFR)

يهدف تحليل احتياج رأس المال العامل إلى ضمان الاستقرار المالي والتشغيلي لمنصة Spectra AI من خلال تحقيق التوازن بين التدفقات النقدية الداخلة والخارجة أثناء دورة النشاط.

التدفقات النقدية الخارجة

تشمل النفقات التشغيلية اللازمة لاستمرار عمل المنصة، مثل:

تكاليف البحث والتطوير.

صيانة البرمجيات وتحديثها.

خدمات الاستضافة السحابية.

رسوم واجهات البرمجة والتكامل مع الذكاء الاصطناعي.

تكاليف التسويق والدعم الفني.

التدفقات النقدية الداخلة

تشمل الإيرادات المتأتية من:

الاشتراكات الشهرية والسنوية.

تراخيص المؤسسات والمختبرات.

خدمات الدفع مقابل التحليل.

الخدمات الإضافية ذات القيمة المضافة.

ويُسهم هذا التحليل في ضمان توفر السيولة اللازمة لتغطية الالتزامات المالية قصيرة الأجل، وإدارة حسابات الموردين ومقدمي الخدمات من جهة، ومستحقات العملاء والاشتراكات من جهة أخرى، بما يضمن استمرارية تشغيل المنصة دون انقطاع والحفاظ على جودة الخدمة المقدمة للمستخدمين.

4. خطة الخزينة:

Month	Projected Revenue (DZD)	Projected Expenses (DZD)	Net Monthly Cash Flow (DZD)
Month 1	[Forecast]	[Dev/Fixed Costs]	[Balance]
Month 2	[Forecast]	[Ops/API Costs]	[Balance]
...	...	...	...
Month 12	[Forecast]	[Marketing/Ops]	[Balance]
Total	Annual Revenue	Total Expenses	Annual Net Profit/Loss

المحور السادس

النموذج الأولي التجريبي

المحور السادس

النموذج الاولي التجريبي

عرض النموذج الأولي لمشروع Spectra AI

الوظائف الأساسية للنموذج الأولي

يُظهر النموذج الأولي لمشروع Spectra AI التكامل الوظيفي بين نظام استقبال ومعالجة البيانات الطيفية ومحرك التحليل المعتمد على الذكاء الاصطناعي. ويتيح هذا التكامل للمستخدم تحميل بيانات FTIR ، وتحليلها تلقائيًا، واستخراج النتائج العلمية في صورة منظمة وقابلة للاستخدام المباشر.

طريقة عرض النموذج الأولي

أ. عرض مرئي (Visual Walkthrough)

يمكن تقديم عرض فيديو مسجل يوضح أهم خصائص المنصة، بما في ذلك:

- مساحة رفع الملفات (Upload Workspace).
- أدوات التحكم في حساسية الكشف عن القمم الطيفية.
- ميزة إنشاء التقرير الأكاديمي بالذكاء الاصطناعي (AI Academic Report).
- خطوات التحليل منذ تحميل البيانات وحتى الحصول على التقرير النهائي.

ب. عرض مباشر وتفاعلي (Live / Interactive Demo)

يمكن عرض المنصة مباشرة أمام لجنة المناقشة من خلال الموقع الإلكتروني:

https://spectradex-ai.lovable.app/

مع تنفيذ تحليل فعلي لملف FTIR تجريبي في الزمن الحقيقي، لإبراز سرعة المعالجة وكفاءة النظام ودقة النتائج.

ج. لقطات شاشة (Screen Captures)

يُنصح بإدراج صور عالية الجودة ضمن شرائح العرض التقديمي توضح:

- واجهة عرض الطيف (Spectrum Plot Workspace).
- واجهة التقرير الأكاديمي المولد بالذكاء الاصطناعي (AI Academic Report Workspace).

وذلك لتمكين أعضاء اللجنة من الاطلاع على تصميم المنصة وتجربة المستخدم حتى في حال عدم توفر اتصال مباشر بالإنترنت أثناء العرض.

خارطة طريق التطوير (الوصول إلى النموذج الأولي)

تم الوصول إلى النموذج الأولي الحالي من خلال مجموعة من المراحل الأساسية المتتابعة:

1. تصميم البنية المعمارية للنظام

وتمثلت في:

- **تصميم الإطار العام للمساعد الذكي لتحليل أطياف** FTIR.
- **تحديد المتطلبات الوظيفية والتقنية للنظام.**
- **وضع هيكلية التكامل بين مكونات التحليل ومعالجة البيانات والذكاء الاصطناعي.**

2. تطوير الخوارزميات

تم تطوير خوارزمية قوية وقابلة للتعديل للكشف عن القمم الطيفية باستخدام لغة Python **والاستفادة من مكتبات علمية متخصصة مثل:**

- Pandas
- NumPy
- SciPy

مما مكّن من توفير تحكم دقيق في حساسية الكشف وتحسين جودة التحليل.

3. هندسة الموجهات (Prompt Engineering)

تم تصميم وتحسين منهجية التواصل مع واجهة Google Gemini API **لضمان:**

- **تفسير صحيح للبيانات الطيفية.**
- **توليد نصوص علمية دقيقة.**
- **إنتاج تقارير متوافقة مع أسلوب الكتابة الأكاديمية ومتطلبات المجلات العلمية.**

4. تطوير واجهة المستخدم

تم تصميم واجهة رسومية حديثة وسهلة الاستخدام توفر:

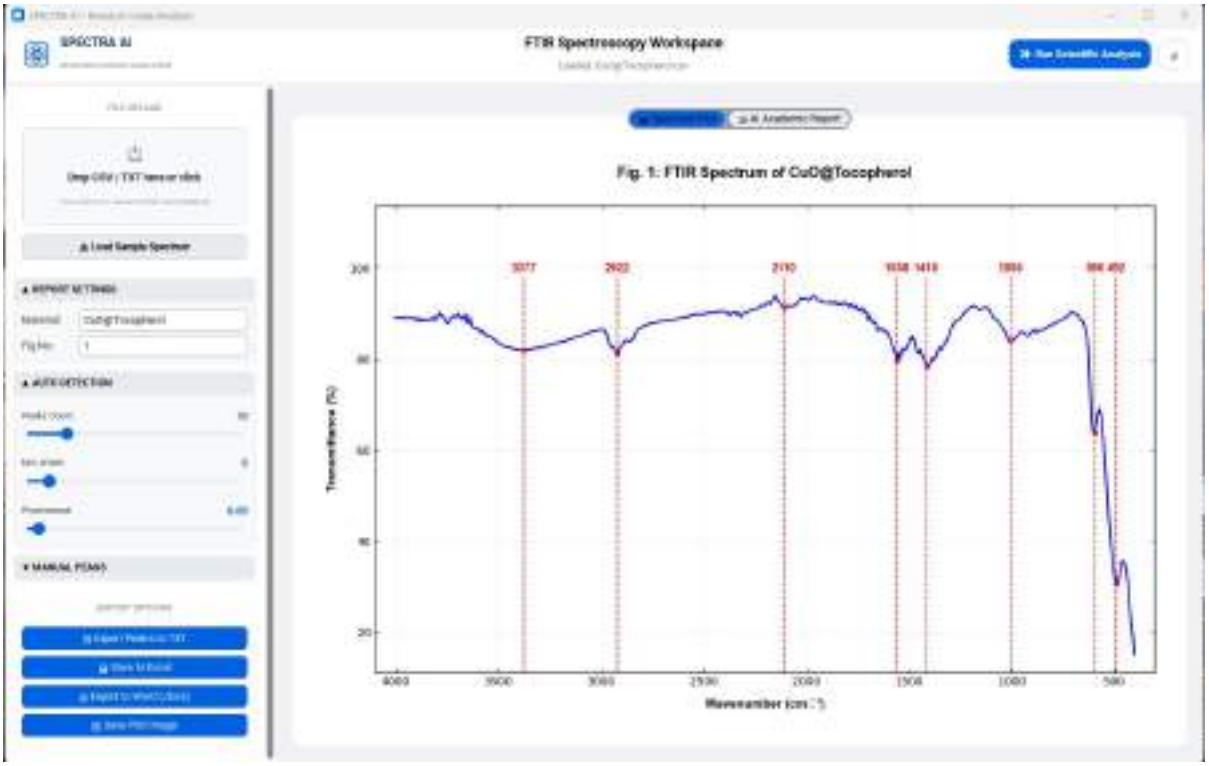
- **بيئة عمل تفاعلية ومبسطة.**
- **سهولة الوصول إلى جميع أدوات التحليل.**
- **إمكانية الاستخدام من قبل الباحثين غير المتخصصين في التحليل الطيفي أو البرمجة.**

5. التحقق والتقييم التجريبي

تم التحقق من موثوقية مخرجات النظام من خلال مقارنة نتائجه مع بيانات FTIR منشورة في الأدبيات العلمية لثلاثة أنظمة مختلفة من المواد النانوية:

- Ag/AgCl
- CuO@Tocopherol
- ZnO/BaMg2

وقد ساهمت هذه المرحلة في التأكد من الدقة العلمية للمنصة وضمان توافق نتائجها مع النتائج المرجعية المنشورة، مما يعزز موثوقيتها للاستخدام الأكاديمي والصناعي.



قائمة الملاحق

الملحق رقم 01

ميزانية المؤسسة الناشئة

BILANS DE STARTUP : Spectra AI

	REALIZED			FORECAST				
Items (In thousands DZD)	N -2	N -1	N (Launch)	N+1	N+2	N+3	N+4	N+5
I. ASSETS								
NON-CURRENT ASSETS	0	0	800	640	480	700	500	1300
(AI Algos & Intangible Assets (Platform	0	0	500	400	300	200	100	500
Tangible Assets (IT Equipment/Servers)	0	0	300	240	180	500	400	800
Other Non-Current Assets	0	0	0	0	0	0	0	0
CURRENT ASSETS	0	0	200	1339	5099	13079	28479	52879
(Subs & Trade Receivables (Customers	0	0	0	150	400	800	1500	2500
Bank Balances & Cash in Hand	0	0	200	1189	4699	12279	26979	50379
TOTAL ASSETS (A)	0	0	1000	1979	5579	13779	28979	54179
LIABILITIES & II. EQUITY								
EQUITY	0	0	1000	1879	5379	13379	28379	53379
Issued Capital	0	0	1000	1000	1000	1000	1000	1000
Retained Earnings	0	0	0	0	879	4379	12379	27379
Net Income	0	0	0	879	3500	8000	15000	25000
NON-CURRENT LIABILITIES	0	0	0	0	0	0	0	0
Financial Debts & Borrowings	0	0	0	0	0	0	0	0
CURRENT LIABILITIES	0	0	0	100	200	400	600	800
(Suppliers & Trade Payables (Cloud	0	0	0	100	200	400	600	800
Taxes & Other Debts	0	0	0	0	0	0	0	0
LIABILITIES (B) & TOTAL EQUITY	0	0	1000	1979	5579	13779	28979	54179
BALANCE CHECK (A - B = 0)	0	0	0	0	0	0	0	0

الملحق رقم 02

جدول حسابات النتائج المتوقعة

COMPTE DE RUSULTAT PREVISIONNEL DE STARTUP : Spectra AI

	REALIZED			FORECAST				
Items (In Thousands DZD)	N -2	N -1	N	N+1	+2	N+3	N+4	N+5
Sales & Related Revenues (SaaS Subscriptions)	0	0	0	3 789	9 160	17 400	29 650	47 900
Change in Inventory of Finished Goods & WIP	0	0	0	0	0	0	0	0
Capitalized Production	0	0	0	0	0	0	0	0
Operating Subsidies	0	0	0	0	0	0	0	0
Total Production for the Year	0	0	0	3 789	9 160	17 400	29 650	47 900
Consumables & Raw Materials	0	0	0	0	0	0	0	0
External Services (Cloud, APIs, Marketing)	0	0	0	1 500	3 000	5 000	8 000	12 000
Total Consumption for the Year	0	0	0	1 500	3 000	5 000	8 000	12 000
Operating Value Added	0	0	0	2 289	6 160	12 400	21 650	35 900
Personnel Costs	0	0	0	1 200	2 400	4 000	6 000	10 000
Taxes & Duties	0	0	0	50	100	200	400	600
Gross Operating Surplus (EBITDA)	0	0	0	1 039	3 660	8 200	15 250	25 300
Other Operating Income	0	0	0	0	0	0	0	0
Other Operating Expenses	0	0	0	0	0	0	0	0

Depreciation, Amortization & Provisions	0	0	0	160	160	200	250	300
Reversals of Impairment & Provisions	0	0	0	0	0	0	0	0
Operating Income (EBIT)	0	0	0	879	3 500	8 000	15 000	25 000
Financial Income	0	0	0	0	0	0	0	0
Financial Expenses	0	0	0	0	0	0	0	0
Financial Result	0	0	0	0	0	0	0	0
Ordinary Income Before Tax (EBT)	0	0	0	879	3 500	8 000	15 000	25 000
Corporate Income Tax (Startup Exemption)	0	0	0	0	0	0	0	0
Deferred Taxes	0	0	0	0	0	0	0	0
Total Revenues from Ordinary Activities	0	0	0	3 789	9 160	17 400	29 650	47 900
Total Expenses from Ordinary Activities	0	0	0	2 910	5 660	9 400	14 650	22 900
Net Income from Ordinary Activities	0	0	0	879	3 500	8 000	15 000	25 000
Extraordinary Income	0	0	0	0	0	0	0	0
Extraordinary Expenses	0	0	0	0	0	0	0	0
Extraordinary Result	0	0	0	0	0	0	0	0
NET INCOME FOR THE YEAR	**0**	**0**	**0**	**879**	**3 500**	**8 000**	**15 000**	**25 000**

الملحق رقم 03

حسابات الخزينة

STARTUP : Spectra AI

	REALIZED			FORECAST				
Items (In Thousands DZD)	N-2	N-1	N	N+1	N+2	N+3	N+4	N+5
CASH FLOWS FROM OPERATING ACTIVITIES								
Net Income for the Year	0	0	0	879	3 500	8 000	15 000	25 000
Adjustments for:								
- Depreciation & Amortization	0	0	0	160	160	200	250	300
- Change in Deferred Taxes	0	0	0	0	0	0	0	0
- Change in Inventory	0	0	0	0	0	0	0	0
- Change in Trade Receivables (Outflow)	0	0	0	-150	-250	-400	-700	-1 000
- Change in Trade Payables (Inflow)	0	0	0	100	100	200	200	200
- Capital Gains/Losses on Disposals	0	0	0	0	0	0	0	0
Net Cash from Operating Activities (A)	0	0	0	989	3 510	8 000	14 750	24 500
CASH FLOWS FROM INVESTING ACTIVITIES								
Capital Expenditures (Software & Servers)	0	0	-800	0	0	-420	-50	-1 100
Proceeds from Sale of Fixed Assets	0	0	0	0	0	0	0	0
Impact of Changes in Consolidation Scope	0	0	0	0	0	0	0	0
Net Cash from Investing Activities (B)	0	0	-800	0	0	-420	-50	-1 100
CASH FLOWS FROM FINANCING ACTIVITIES								
Dividends Paid to Shareholders	0	0	0	0	0	0	0	0
Capital Increase (ASF - Startup Fund Share)	0	0	500	0	0	0	0	0
Capital Increase (Founder Share)	0	0	500	0	0	0	0	0
ASF Current Account Injection	0	0	0	0	0	0	0	0
Repayment of ASF Capital (Nominal Value)	0	0	0	0	0	0	0	0
Repayment of ASF Current Account	0	0	0	0	0	0	0	0
Net Cash from Financing Activities (C)	0	0	1 000	0	0	0	0	0
Net Change in Cash for Period (A+B+C)	0	0	200	989	3 510	7 580	14 700	23 400
Opening Cash Balance (Start of Period)	0	0	0	200	1 189	4 699	12 279	26 979
Closing Cash Balance (End of Period)	0	0	200	1 189	4 699	12 279	26 979	50 379
Net Change in Cash (Validation Check)	0	0	200	989	3 510	7 580	14 700	23 400

الملحق رقم 04

نموذج العمل التجاري

شرائح العملاء	العلاقات مع العملاء	القيم المقترحة	الأنشطة الرئيسية	الشراكات الرئيسية
• الباحثون الأكاديميون (جامعات ومخابر البحث) • طلبة الدراسات العليا (ماجستير / دكتوراه) • المخابر الصناعية (كيمياء، مواد أولية، طاقة) • المخابر الطبية (تحاليل متقدمة)	• دعم تقني عبر البريد الإلكتروني أو الدردشة • محتوى تعليمي فيديوهات تعليمية • تجربة مجانية • دعم مخصص للمؤسسات	• تحليل سريع ودقيق للطيف الأشعة تحت الحمراء باستخدام الذكاء الاصطناعي • توليد تقارير علمية جاهزة للنشر تقليل الوقت والجهد مقارنة بالتحليل اليدوي • دعم تحديد المجموعات الكيميائية الوظيفية • واجهة سهلة الاستخدام بدون الحاجة لخبرة عميقة • تقليل الأخطاء البشرية في تفسير الأطياف • إمكانية الاستخدام عبر الإنترنت بدون تثبيت برامج	• تطوير وتحسين نموذج الذكاء الاصطناعي تحديث قاعدة البيانات الطيفية •صيانة المنصة وتحسين الأداء •التسويق الرقمي واستقطاب المستخدمين •دعم العملاء •ضمان جودة النتائج العلمية	• مزودي خدمات الحوسبة السحابية • باحثين في الكيمياء التحليلية • جامعات ومراكز بحث
القنوات			**الموارد الرئيسية**	
منصة ويب رئيسية (SaaS) حملات تسويق عبر : □ رسائل التواصل الاجتماعي (فايسبوك – انستغرام ...) □ ResearchGate و LinkedIn الباحثين • شراكات مع جامعات ومخابر • مؤتمرات علمية ومعارض • ورش تدريب : (Workshops / Webinars)			• خوارزميات الذكاء الاصطناعي لتحليل الأشعة تحت الحمراء • قاعدة بيانات أطياف مرجعية • البنية التحتية السحابية (Cloud) • فريق تقني (ذكاء اصطناعي + تطوير) • خبرة علمية في الكيمياء التحليلية	

مصادر الإيرادات	هيكل التكاليف
• الاشتراكات شهرية وسنوية: باقة الطلبة (منخفضة السعر) - باقة الباحثين - باقة المؤسسات / الشركات • الدفع حسب الاستخدام (Pay-per-analysis) • تراخيص للمخابر (Enterprise Licensing) • خدمات إضافية : تقارير مفصلة و مخصصة	• تطوير البرمجيات والذكاء الاصطناعي • استضافة وخدمات سحابية • التسويق والإعلانات • صيانة وتحديث قاعدة البيانات

دليل مشروع

للحصول على شهادة مؤسسة ناشئة
في إطار القرار الوزاري 1275

ديسمبر
2022

CNCSIIU — MHESR — ALGERIA