JUPITER PRIME

Consulting • Training • Advisory

جوبيتر برايم للاستشارات والتدريب

COURSE BROCHURE

# Advanced ChatGPT Multimodal AI Integration Training Course

12 – 23 July 2027
New York

**jupiterprime.uk**

info@jupiterprime.uk

Registered in the United Kingdom and Saudi Arabia

# Advanced ChatGPT Multimodal AI Integration Training Course

**Ref:** 22_576 **Date:** 12 – 23 July 2027 **Location:** New York **Fees:** 17000 GBP

ChatGPT Multimodal AI Integration is a 2-week advanced training course for digital product leads, content managers and innovation teams who combine text, images, voice, documents and data inside ChatGPT and leave with a multimodal solution design. Organisations now hold vast stores of scans, photos and recordings that text-only assistants ignore, and vision and speech models finally make them usable. Jupiter Prime Training runs this course for teams turning mixed media into reliable business answers.

## Why ChatGPT Multimodal AI Integration Matters Now

Three forces shape this subject: assistants that see images and hear speech in one conversation, product teams asked to unlock document and media archives, and rising expectations around privacy for faces and voices. An organisation that integrates modalities carefully shortens claims, inspections and content cycles; one that does it carelessly risks misread figures, invented detail and exposure of personal media. The course rests on ISO/IEC 42001, the international standard for an artificial intelligence management system, and the NIST AI Risk Management Framework NIST AI RMF for identifying risks specific to vision and speech.

## Who Gains Most from ChatGPT Multimodal AI Integration

- Leaders who own digital products and decide which features gain image, voice or file inputs
- Managers responsible for content libraries, media assets and publishing quality
- Innovation specialists who test emerging capabilities and shape prototypes
- Owners of document-heavy processes such as claims, onboarding or inspections
- Customer experience designers who shape voice and chat interactions

Product and innovation groups gain most when a product lead, a content owner and a process owner attend together, since each multimodal solution crosses all three areas.

## ChatGPT Multimodal AI Integration Capabilities

Participants leave able to judge, prompt, test and connect multimodal features so that mixed inputs produce answers people can trust.

- Assess which business scenarios benefit from image, voice, document or data inputs — the product owner setting the roadmap
- Design prompts that extract structured fields from scans, photos and recordings — the process owner replacing manual keying
- Test multimodal output against ground-truth samples — quality leads approving release
- Specify privacy and consent controls for images and recordings — data protection and legal reviewers
-

Build integrations that pass media into existing systems — the IT partner maintaining the solution
- Communicate accuracy, cost and latency trade-offs — sponsors funding the next stage

## Key ChatGPT Multimodal AI Integration Terms

- Multimodal model — an AI model that accepts and relates several input types, such as text, images, audio and files, in one request.
- Optical character recognition — technology that converts printed or handwritten text in an image into machine-readable characters.
- Speech-to-text — conversion of spoken audio into written text, the first stage of most voice assistant designs.
- Ground truth — a verified set of correct answers used to measure how accurately a model extracts or describes content.
- Visual hallucination — a model describing objects, text or figures that do not exist in the supplied image.
- Content provenance — information attached to media showing how it was created or altered, including whether AI generated it.

## ChatGPT Multimodal AI Integration Day by Day

### Day 1: Multimodal Capability Landscape

- Modality Map of Text, Image, Voice, Files and Data — Scoping What Each Input Adds
- Vision and Speech Model Limits — Setting Realistic Accuracy Expectations
- Multimodal Use Case Canvas — Linking Inputs to Business Outcomes
- Sector Scan from Insurance Claims to Retail Merchandising
- Use Case Prioritisation Grid — Choosing Candidates Worth Prototyping

### Day 2: Prompt Design Across Modalities

- Image Prompt Framing — Directing Attention to Regions and Attributes
- Combined Text and Image Instructions — Resolving Conflicting Signals
- Structured Extraction Templates — Turning Visual Content into Fields
- Reference Image Sets — Anchoring Consistent Visual Judgements
- Modality-Specific Failure Patterns — Misreads, Omissions and Invented Detail

### Day 3: Document Understanding

- Scanned Form and Invoice Parsing — Capturing Fields Without Manual Keying
- Table and Chart Interpretation — Reading Figures from Reports
- Long PDF Navigation — Chunking, Citing Pages and Cross-Checking
- Handwriting and Low-Quality Scan Handling — Deciding When to Escalate
- Document Classification Rules — Routing Files to the Right Process

## Day 4: Visual Inspection and Field Notes

- Photo-Based Inspection Notes — Describing Condition, Defects and Severity
- Annotated Image Workflows — Marking Evidence for Reviewers
- Before-and-After Comparison Prompts — Detecting Change Between Images
- Site Photo Logs in Facilities and Construction Settings
- Confidence Flags — Separating Clear Findings from Uncertain Ones

## Day 5: Voice Assistants and Audio

- Speech-to-Text Transcription Quality — Accents, Noise and Terminology
- Voice Mode Conversation Design — Turn-Taking, Brevity and Confirmation
- Meeting and Call Summaries — Extracting Actions and Owners
- Text-to-Speech Output Choices — Tone and Accessibility Considerations
- Bilingual Voice Interactions — Handling Arabic and English Switching

## Day 6: Data Files and Advanced Analysis

- Spreadsheet Upload Analysis — Profiling, Cleaning and Summarising Datasets
- Chart Generation from Data — Choosing Visuals for Decision Makers
- Combining Data with Images and Documents — Building Evidence Packs
- Reproducibility Notes — Recording Steps Behind Every Generated Result
- Data Sensitivity Screening — Deciding What Files May Be Uploaded

## Day 7: Quality Checks for Multimodal Output

- Ground-Truth Sample Sets — Measuring Extraction and Description Accuracy
- Human Review Sampling Plans — Sizing Checks to Risk Level
- Hallucination Detection in Visual Answers — Spotting Invented Objects
- Bias Checks Across Images and Voices — Testing Diverse Inputs
- Quality Scorecards — Reporting Accuracy Trends to Product Owners

## Day 8: Integration into Existing Tools

- Multimodal API Calls — Sending Images, Audio and Files Programmatically
- Connecting to Content Management and Ticketing Systems
- Custom GPTs with Vision and File Knowledge — Packaging Repeatable Tasks
- Latency, File Size and Cost Trade-Offs — Choosing Processing Routes
- Fallback Design — Keeping Processes Running When Recognition Fails

## Day 9: Responsible Multimodal Use

- Privacy in Images and Recordings — Faces, Voices and Personal Details
- Consent and Disclosure Practices for Voice and Camera Capture
- ISO/IEC 42001 Controls — Managing Multimodal Systems Within an AI Management System
- NIST AI RMF Mapping — Identifying Risks Specific to Vision and Speech
- Content Provenance and Labelling — Marking AI-Generated Media

## Day 10: Practising the Multimodal Solution Design

- Use Case Prioritisation Applied — Selecting a Real Scenario from Participants' Teams
- Cross-Modality Prompt Set — Assembling Extraction and Inspection Prompts
- Quality Scorecard Trial — Testing Accuracy Against a Sample Set
- Integration and Fallback Sketch — Placing the Solution in Existing Tools
- Multimodal Solution Design — Completing and Presenting the Final Document

## Applying ChatGPT Multimodal AI Integration After the Course

- Run a ground-truth accuracy test on one document or image process and publish a quality scorecard for its owner
- Prototype one voice or photo-based assistant with defined fallback paths and privacy controls agreed with data protection reviewers

## ChatGPT Multimodal AI Integration Questions FAQ

### What is ChatGPT multimodal AI integration, and why does it matter now?

It is the design of business solutions in which ChatGPT reads images, documents, audio and data together. It matters now because vision and speech capabilities have matured, letting organisations use scans, photos and recordings that previously needed manual handling.

### Which roles gain most from ChatGPT multimodal AI integration?

Digital product leads, content managers and innovation teams gain most. You decide which features accept media, own the quality of content or test new capabilities, so you can turn mixed inputs into dependable services.

### How does ChatGPT multimodal AI integration training differ from a general ChatGPT productivity course?

It focuses on images, voice, documents and data rather than text prompts alone. A productivity course improves everyday writing tasks, whereas this subject covers extraction, inspection notes, voice design, accuracy testing and system integration.

### What do you bring back from ChatGPT multimodal AI integration training?

You bring back a solution design for one real multimodal scenario. It includes prompts for each input type, a quality scorecard, integration and fallback routes, and privacy controls ready for review.

## ChatGPT Multimodal AI Integration: The Way Forward

The multimodal solution design first informs which scenario moves to a funded prototype and which input types it accepts. Product leads, content owners and IT partners use it to agree prompts, accuracy thresholds and integration routes. After the first live month, revisit the ground-truth samples and fallback paths, because real photos, scans and recordings vary far more than prepared examples.