



DIGITAL INNOVATION AND TRANSFORMATION

ChatGPT Advanced: Mastering Multimodal AI Integration

Course code: DIT-020

SPARK TRAINING

Professional training courses for leaders, specialists, and organizations seeking practical capability development.

info@sparktraining.eu

+31 85 115 0004 | sparktraining.eu

— Workshop Overview

Artificial intelligence is rapidly evolving beyond text-based interactions into multimodal systems capable of processing and generating information across text, images, audio, video, and structured data. These developments are creating new opportunities for organisations to improve productivity, automate complex workflows, enhance customer experiences, and develop innovative digital services. ChatGPT and related AI technologies enable professionals to interact with information through natural language while incorporating visual understanding, document analysis, voice interactions, data interpretation, and integrated business workflows. However, realising the full value of multimodal AI requires more than familiarity with conversational prompting. Organisations must understand how different AI modalities work together, how to design effective multimodal prompts, how to integrate AI capabilities with enterprise systems, and how to evaluate the accuracy, reliability, security, and business value of AI-generated outputs. Advanced applications also introduce challenges involving inconsistent model responses, hallucinations, sensitive information, system integration, computational costs, accessibility, and responsible AI governance. Successful implementation therefore requires a structured approach combining technical understanding, practical experimentation, workflow design, and organisational controls. The ChatGPT Advanced: Mastering Multimodal AI Integration course provides participants with practical knowledge of advanced ChatGPT capabilities and multimodal AI integration techniques. Participants explore multimodal prompting, image and document understanding, speech-based interactions, video analysis concepts, structured data processing, API integration, workflow automation, and the development of AI-enabled business applications. Through practical demonstrations, guided exercises, and applied implementation workshops, participants develop the ability to design, evaluate, and deploy multimodal AI solutions that address real organisational challenges.

Why Attend

The programme combines instructor-led discussions, live demonstrations, hands-on ChatGPT exercises, multimodal prompting workshops, integration design activities, and practical AI application development scenarios. Participants work with realistic business materials, including documents, spreadsheets, images,

Who Should Attend

Digital transformation managers and specialists
Artificial intelligence and innovation professionals
IT managers and enterprise technology specialists
Business analysts and process improvement professionals
Software developers and application integration specialists

— Learning Objectives

- Explain the principles and capabilities of multimodal generative AI systems

- Identify appropriate business applications for text, image, audio, video, and structured data processing
- Develop advanced prompts combining multiple information modalities
- Apply ChatGPT to complex document and visual information analysis
- Design effective speech-based and conversational AI interactions
- Understand video interpretation workflows and their technical limitations
- Integrate multimodal AI capabilities with business applications through APIs and automation tools
- Develop structured AI workflows using external information and business data
- Evaluate multimodal AI outputs for accuracy, consistency, relevance, and reliability
- Apply privacy, security, accessibility, and responsible AI principles
- Assess the operational and financial value of multimodal AI applications
- Design a practical multimodal AI integration solution for an organisational use case

Course Outline

**DAY
1**

Advanced ChatGPT Capabilities and Multimodal AI Foundations

- Understanding the evolution of generative AI from text-based language models to multimodal systems that process text, images, audio, video, and structured information
- Exploring the capabilities, limitations, and practical differences between multimodal input processing, reasoning, and content generation
- Examining advanced ChatGPT features, model selection considerations, context management, and available multimodal interaction methods
- Developing advanced prompting strategies using role definition, task decomposition, contextual information, examples, output constraints, and iterative refinement
- Understanding how to combine textual instructions with images, documents, tables, and other inputs to improve task execution and contextual interpretation
- Practical workshop: Designing and testing advanced multimodal prompts for business analysis, visual interpretation, document review, and structured information extraction

**DAY
2**

Image Intelligence, Document Understanding and Visual AI Applications

- Applying multimodal AI to interpret images, diagrams, charts, screenshots, and visual business information
- Analysing complex documents containing text, tables, figures, financial information, and mixed visual layouts
- Designing document intelligence workflows for classification, summarisation, information extraction, comparison, and question answering
- Exploring image generation and editing workflows, including prompt design, creative iteration, visual consistency, and quality evaluation
- Understanding visual AI limitations involving image resolution, ambiguous content, numerical accuracy, document structure, and unsupported visual assumptions
- Hands-on exercise: Developing a multimodal document analysis workflow that extracts information, evaluates visual content, generates structured findings, and applies human verification

**DAY
3**

Audio, Voice, Video and Real-Time Multimodal Interactions

- Understanding speech recognition, speech synthesis, audio understanding, and conversational AI capabilities within multimodal systems
- Designing voice-enabled workflows for customer interactions, meeting assistance, language support, accessibility, and internal knowledge retrieval
- Exploring audio transcription, speaker-related considerations, summarisation, and extraction of action items from spoken information

- Understanding video analysis concepts, including frame interpretation, temporal context, scene understanding, and the challenges of processing extended video sequences
- Examining real-time multimodal interaction architectures, latency considerations, session management, user experience, and error handling
- Practical workshop: Designing a voice-enabled assistant and a video or audio analysis workflow with clear inputs, outputs, validation procedures, and operational safeguards

DAY
4

Multimodal AI Integration, APIs and Intelligent Workflow Automation

- Understanding multimodal AI integration architecture, including application interfaces, API requests, model inputs, structured responses, and external system connections

- Exploring the use of OpenAI APIs and relevant development tools to process supported combinations of text, images, audio, and other media
- Designing workflows that connect AI services with document repositories, business applications, databases, and automation platforms
- Applying structured outputs, function calling concepts, retrieval-augmented generation, and workflow orchestration to support reliable AI-enabled processes
- Evaluating integration challenges involving authentication, access permissions, data transfer, error handling, latency, scalability, and operational costs
- Integration workshop: Designing an end-to-end multimodal AI workflow that receives business information, processes multiple input types, produces structured outputs, and routes results for human review

DAY
5

Responsible Multimodal AI, Performance Evaluation and Enterprise Deployment

- Establishing evaluation methods for multimodal AI applications covering accuracy, relevance, completeness, consistency, latency, usability, and business effectiveness
- Identifying multimodal AI risks involving hallucinations, prompt injection, misleading visual interpretation, sensitive information exposure, intellectual property, and inappropriate automation
- Applying responsible AI principles through human oversight, access controls, data minimisation, output validation, transparency, and documented governance procedures
- Developing deployment strategies covering pilot selection, technical readiness, stakeholder engagement, user training, operational monitoring, and continuous improvement
- Assessing multimodal AI business cases through implementation costs, productivity improvements, quality indicators, risk reduction, and return on investment
- Capstone exercise: Developing and presenting a Comprehensive Multimodal AI Integration Solution incorporating a defined business use case, advanced prompting strategy, multimodal processing workflow, integration architecture, evaluation criteria, security controls, and implementation roadmap

— Upcoming Classroom Dates

Schedule available on request. Contact SPARK Training for current dates, locations, and fees.

— Certificate and Registration

Participants who attend and complete the full programme receive a SPARK Training Certificate of Completion. For registration, in-house delivery, or availability questions, contact info@sparktraining.eu or +31 85 115 0004.