JUPITER PRIME

Consulting • Training • Advisory

جوبيتر برايم للاستشارات والتدريب

COURSE BROCHURE

Measuring LLM and AI Agent Productivity Gains Training Course

jupiterprime.uk

info@jupiterprime.uk

Registered in the United Kingdom and Saudi Arabia

Measuring LLM and AI Agent Productivity Gains Training Course

Measuring LLM and AI Agent Productivity Gains is a 1-week training course for IT leads, digital managers and business analysts who leave with a productivity measurement scorecard for large language model LLM assistants, retrieval-augmented generation RAG services and AI agent workflows. Boards fund these deployments for saved hours, yet few can show evidence. Jupiter Prime Training delivers this course on measuring LLM and AI agent productivity gains.

Why Measuring LLM and AI Agent Productivity Gains Matters Now

Three pressures meet. Licence and compute costs for enterprise LLM services grow with every seat and token, so finance functions ask for proof of return. Agent workflows now act across ticketing, document and email systems, which makes errors costlier and attribution harder. Uneven adoption means averages hide teams that lose time correcting output. Organisations that baseline carefully can redirect spend towards the use cases that pay back; those that rely on anecdote fund the loudest pilots. The course anchors measurement to ISO/IEC 42001 Artificial Intelligence Management System and the NIST AI Risk Management Framework NIST AI RMF.

Who Gains Most from Measuring LLM and AI Agent Productivity Gains

- IT leads who own the LLM platform, RAG pipelines and usage telemetry
- Digital managers who prioritise assistant and agent use cases across departments
- Business analysts who design baselines, sampling and before-and-after comparisons
- Process owners whose teams use assistants for drafting, search or case handling
- Data stewards who set retention, access and quality rules for retrieval sources

A digital transformation office gains most when its platform, analysis and process members attend together and agree one measurement method.

Measuring LLM and AI Agent Productivity Gains Capabilities

Participants practise separating genuine gains from novelty effects, rework and shadow usage.

- Design task-level baselines before an assistant or agent goes live — funding decisions
- Calculate time saved, rework rate and cost per resolved task — finance partners
- Test RAG answer quality with groundedness and citation accuracy checks — knowledge owners
- Specify agent workflow controls, escalation points and audit logs — risk and IT security
- Weigh usage telemetry against survey evidence of perceived productivity — sponsors
- Communicate a scaling or stopping recommendation from measured results — executive committee

Key Measuring LLM and AI Agent Productivity Gains Terms

- Large language model LLM — A model trained on large text collections that generates and summarises language in response to prompts.
- Retrieval-augmented generation RAG — A method that fetches approved documents and passes them to an LLM so answers draw on organisational sources.
- AI agent — Software that plans and executes multi-step tasks by calling tools and systems with limited human input.
- Productivity baseline — Measured time, cost and quality of a task before an assistant is introduced, used for comparison.
- Groundedness — The degree to which a generated answer is supported by the retrieved source passages.
- Rework rate — The share of AI-generated outputs that people must correct or redo before use.

Measuring LLM and AI Agent Productivity Gains Day by Day

Day 1: Where LLM Value Is Claimed

- LLM, RAG and Agent Capability Map — Which Tasks Each Can Change
- Use Case Inventory by Function — Separating Drafting, Search and Action
- Hype Versus Evidence Review — Questioning Vendor Productivity Claims
- ISO/IEC 42001 and NIST AI RMF Measurement Clauses — Accountability for Results
- Token and Licence Cost Model — Total Spend per Active User

Day 2: Baselines and Experiment Design

- Task Time and Motion Sampling — Establishing Pre-Deployment Baselines
- Control and Treatment Group Design — Isolating Assistant Effects
- Quality Scoring Rubrics — Accuracy, Completeness and Tone of Outputs
- Usage Telemetry Extraction — Prompts, Sessions and Active Seats
- Perceived Productivity Surveys — Reconciling Opinion With Logged Behaviour

Day 3: Measuring RAG and Agent Workflows

- RAG Evaluation Sets — Groundedness and Citation Accuracy Scoring
- Retrieval Source Data Rules — Access, Freshness and Retention Decisions
- Agent Task Completion Logs — Success, Escalation and Failure Rates
- Human-in-the-Loop Checkpoints — Where Approval Protects Quality
- Rework and Error Cost Tracking — Hidden Losses Behind Fast Output

Day 4: From Metrics to Investment Decisions

- Cost per Resolved Task Calculation — Comparing Assistants With Current Methods
- Benefit Attribution Methods — Crediting Gains Fairly Across Teams
- Scale, Fix or Retire Matrix — Deciding the Fate of Each Use Case
- Risk-Adjusted Return View — Weighing Errors Against Time Saved
- Executive Reporting Narrative — Presenting Uncertainty Honestly

Day 5: Practising the Productivity Measurement Scorecard

- Baseline Rehearsal on a Live Use Case — Applying Day 2 Sampling
- RAG and Agent Metric Selection — Applying Day 3 Evaluation Sets
- Investment Case Drafting — Applying Day 4 Cost per Task
- Peer Challenge Panel — Testing Assumptions and Attribution Logic
- Productivity Measurement Scorecard — Finalising Indicators, Owners and Review Cadence

Applying Measuring LLM and AI Agent Productivity Gains After the Course

- Run task sampling on one assistant use case to set a baseline before rollout
- Build a RAG evaluation set of real employee questions and score groundedness monthly

Measuring LLM and AI Agent Productivity Gains Questions FAQ

What is measuring LLM and AI agent productivity gains, and why does it matter now?

It is the practice of proving, with baselines and evidence, how much time and quality LLM assistants, RAG services and agents add. It matters now because spending is rising faster than proof of return.

Which roles gain most from measuring LLM and AI agent productivity gains?

IT leads, digital managers and business analysts gain most. Each controls part of the evidence chain, from telemetry to investment cases.

How does measuring LLM and AI agent productivity gains differ from a course on building RAG assistants?

It concentrates on evidence and decisions rather than construction. A build-focused course designs retrieval pipelines; this one measures whether those pipelines save time in settings such as insurance claims handling and telecom customer service.

What do you bring back from measuring LLM and AI agent productivity gains?

You bring back a ready scorecard with baselines, quality checks, cost per task and named owners for one real use case.

Measuring LLM and AI Agent Productivity Gains: The Way Forward

The productivity measurement scorecard first informs whether a live assistant or agent use case should scale, be fixed or retire. Digital managers, finance partners and process owners use it at each quarterly portfolio review. After its first use, the perspective worth revisiting is attribution as several assistants touch the same task.