

Working Machines: Beyond Automation

Glossary

This glossary defines the key terms, concepts and frameworks used throughout Working Machines: Beyond Automation, including several terms coined or given particular meaning in the book. Where an entry relates closely to a specific discussion, the relevant chapter is indicated in brackets. A list of abbreviations follows the main glossary.

A B C D E F G H I K L M N O P R S T U V W

Abbreviations

A

Agentic AI. AI systems that can autonomously plan, execute and monitor complex, multi-step tasks by breaking large goals into sub-tasks, using digital tools and self-correcting along the way. Unlike traditional AI, which responds to a single prompt, agentic systems pursue objectives with minimal human input. A central theme of Part 3: see also Autonomous agent.

AI-fuelled impostor syndrome. Workers doubting their own skills and value because of their reliance on AI tools. Reported by 27% of UK workers in Ivanti's 2025 Technology at Work research (Chapter 4).

AI Ghostwriter Effect. The finding (Draxler et al., 2024) that users of AI-generated text often do not consider themselves the authors of the work, even when they have made significant revisions to it. Central to the discussion of authorship in Chapter 4.

AI governance. The structures, processes and accountability mechanisms through which organisations and governments direct how AI is designed, deployed and audited. Mature governance is distributed across functions rather than confined to IT or legal (Chapters 5 and 6).

AI literacy. The ability to understand what AI can and cannot do, and to interrogate its outputs with the same rigour applied to a human advisor. Now recognised as a core executive and workforce skillset (Chapter 1).

AI natives. Gen Z and Gen Alpha: the generations for whom consulting an AI model is the default starting point rather than a novelty. Their challenge is differentiation and independent thought, not adoption (Chapters 4 and 11).

AI persona development. An emerging discipline combining tone design, ethics framing and interaction scripting to shape how an AI system introduces itself, handles difficult queries, frames uncertainty and builds trust (Chapter 3).

Algorithm. A defined set of rules or instructions a computer follows to solve a problem or complete a task. Machine learning algorithms adjust their own behaviour

based on patterns in data.

Algorithmic bias. Systematic unfairness in AI outputs, usually inherited from skewed training data, unrepresentative development teams or flawed design. Examples in the book include Amazon's recruiting tool penalising women's CVs and racially divergent psychiatric treatment recommendations (Chapter 6).

Algorithmic erasure. The silencing of entire communities when moderation and ranking systems, built primarily for majority languages, disproportionately remove or downrank their content. Illustrated by AI bias against Indian regional-language creators (Chapter 6).

Alignment. The task of making an AI system's behaviour match human intentions, values and safety expectations. Techniques include Reinforcement Learning from Human Feedback and Constitutional AI.

Annotation (data labelling). The human work of tagging text, images, code and other data so machine learning systems can learn from it. Often outsourced to low-wage workers under tight guidelines: see Ghost work (Chapter 9).

API aggregators. Middleware companies (for example, LangChain and Pinecone) whose tools stitch together multiple AI services, such as memory, vector search and chaining logic, into user-ready applications. Quietly influential through standard-setting and dependency creation (Chapter 10).

API (Application Programming Interface). A defined way for software systems to talk to each other. Most businesses access AI models through APIs, which makes them exposed to sudden pricing and policy changes by providers (Chapter 10).

Authorship line. The book's term for the increasingly blurred boundary where human judgment ends and machine generation begins. When the first draft is machine-made, authorship shifts from generation to direction, and accountability can become murky (Chapter 4).

Automated tool. A system that executes a predefined task with minimal variation: reactive, deterministic and always under direct human control. A script that sends monthly invoices is an automated tool. Contrast with Autonomous agent (Chapter 11).

Autonomous agent. A system that pursues goals, makes decisions and adapts its actions on its own. It is proactive, context-aware, non-deterministic and able to resolve unforeseen problems within its domain, which makes accountability harder to assign. Contrast with Automated tool (Chapter 11).

B

Black box. A system whose internal reasoning cannot be inspected or explained, even by its creators. Black-box decision-making is a recurring risk throughout the book, from hiring filters to strategic recommendation engines.

Bounded ethicality. Max Bazerman and Ann Tenbrunsel's concept describing how good people in well-intentioned systems make bad calls without realising it, because their attention is focused elsewhere. Used to explain the Galactica failure (Chapter 9).

C

Chatbot. A software application that conducts conversations with users, typically through text. Modern AI chatbots such as customer-service assistants can handle the large majority of standard queries without human intervention.

Chief Data Officer (CDO). The executive responsible for stewarding data across an organisation's silos, standards and competing agendas. The role has shifted from governance and compliance to strategy and value creation; some organisations use the broader title Chief Data and Analytics Officer (CDAO) (Chapter 5).

Chip sovereignty. National efforts to reduce dependency on concentrated semiconductor supply chains, exemplified by the US CHIPS and Science Act and the European Chips Act (Chapter 12).

Cloud computing. Renting processing power, storage and AI-ready environments from large providers rather than owning the hardware. The book describes AWS, Microsoft Azure and Google Cloud as the 'cloud cartel': the landlords of the AI economy (Chapter 10).

Cobots (collaborative robots). Robots designed to work safely alongside humans in shared spaces, common in manufacturing and logistics (Chapter 2).

Cognitive lift. The offloading of mental load to intelligent automation: summarisation, drafting, transcription, scheduling and other small decisions that drain attention. One of automation's most valuable and least discussed contributions (Chapter 2). Over-reliance carries a cost: see Cognitive offloading.

Cognitive offloading. Delegating thinking tasks to AI to the point that independent reasoning and critical-thinking skills may atrophy. Research cited in Chapter 8 links heavy AI usage with declines in reflective problem-solving.

Command-line tools. Text-based programs that control a computer's core functions. Relevant to AI safety testing, where models given access to such tools can simulate real-world actions (Chapter 7).

Compute. The processing capacity, chiefly high-performance GPUs in data centres, required to train and run AI models. Access to compute is a form of strategic power (Chapters 10 and 12).

Constitutional AI. Anthropic's framework for training models against an explicit set of written principles, giving structure to alignment work that previously ran on slogans (Chapter 6).

Content moderation. The reviewing and filtering of user-generated or AI-generated content. The book highlights both the human cost borne by moderators and the over-blocking that occurs when safety filters cannot distinguish harmful intent from legitimate context (Chapters 6 and 9).

Cross-race effect. The tendency for people, and the AI systems trained on their data, to identify faces and emotions of their own race more accurately than those of others. A key cause of emotion AI's cultural blind spots (Chapter 6).

Crowdsourcing. Drawing on the distributed knowledge of large groups to solve problems or produce work, a concept popularised by Jeff Howe. Underpins the book's case for user-in-the-loop oversight (Chapter 9).

Customer journey mapping. Tracking how customers interact with a company across all touchpoints, increasingly informed by AI-driven analytics (Chapter 5).

D

Data centre. A facility housing the servers that train and run AI systems. Data centres consume vast quantities of electricity and water for cooling, and could account for a significant share of global emissions by 2040 (Chapter 12).

Data colony. The book's term for regions treated as sources of data, minerals and cheap annotation labour without meaningful decision-making power in the AI economy (Chapter 10).

Data ethics. The principles and practices that ensure data is collected, used and protected responsibly (Chapter 5).

Data governance. The distributed practice of keeping data healthy, transparent and trustworthy across an organisation: shared accountability, embedded processes, common language, joint problem-solving, balanced trade-offs and visible executive sponsorship (Chapter 5).

Data lineage. The ability to trace where information comes from and how it is used as it moves through an organisation's systems (Chapter 5).

Data silo. Data held within one department or system and inaccessible to the rest of the organisation. Silos are a primary obstacle to treating data as a strategic asset (Chapter 5).

Deepfake. Synthetic audio, image or video that convincingly depicts a real person doing or saying something they did not. A defining risk of the Synthetic Age (Chapter 13).

Digital divide. The gap between those with reliable access to devices, connectivity, electricity and digital skills and those without. The book argues AI has deepened rather than closed this divide (Chapter 8).

Digital twin (AI twin). A synthetic replica of a person, trained on their footage, voice and style, capable of appearing and speaking on their behalf, such as a CEO's avatar delivering a keynote in twelve languages (Chapter 13).

E

Emergent behaviour. Capabilities or strategies an AI model displays that its developers neither coded nor foresaw: for example, Claude Opus 4's attempts to escalate perceived ethical violations in safety testing, or Bing's 'Sydney' persona (Chapters 7 and 11).

Emotion AI (emotion recognition). Systems that claim to detect emotional states from faces, voices or text. The field is scientifically contested and prone to racial, cultural and neurological bias; Microsoft retired its Azure Face emotion-recognition

functionality in 2022 (Chapter 6).

ESG (Environmental, Social and Governance). The framework investors and boards use to evaluate a company's non-financial performance. AI's energy and water appetite is now flagged as a material ESG risk (Chapters 1 and 12).

Ethics theatre (performative ethics). Visible ethical gestures, such as advisory boards without budgets, principle documents unmoored from practice and transparency reports without metrics, that create the appearance of responsibility without its substance (Chapter 6).

EUV (Extreme Ultraviolet) lithography. The technology, embodied in ASML's roughly US\$200 million machines, required to manufacture the world's most advanced chips. A single point of failure in the AI supply chain (Chapter 10).

E-waste. Discarded electronic hardware. AI's rapid upgrade cycles generate substantial e-waste, often processed in countries with weak recycling safeguards, with rare minerals seldom recovered at scale (Chapter 12).

Explainability. The ability to understand and communicate how an AI system produced a given output. The book argues explainability should be non-negotiable wherever recommendations drive real decisions (Chapter 13).

F

Fab (semiconductor fabrication plant). A factory that manufactures chips. Fabs are among the thirstiest industrial facilities on the planet, and their geographic concentration, with Taiwan producing over 90% of advanced chips, creates strategic fragility (Chapter 12).

Fine-tuning. Further training a foundation model on specific data so it performs better on particular tasks or matches a brand's tone and requirements.

Fluency heuristic. Daniel Kahneman's observation that information which is easy to process and sounds coherent is more readily believed to be true. It explains why confident, polished AI output is so persuasive even when wrong (Chapter 9).

FOMO-led innovation. Investing in AI to signal modernity rather than to solve problems. In a 2024 survey, two-thirds of CEOs admitted fear of missing out drove their AI investments (Chapter 7).

Foundation model. A large AI model, such as those behind ChatGPT or Claude, trained on broad data and adaptable to many downstream tasks through fine-tuning (Chapter 1).

Fourth Industrial Revolution (4IR). The current era of technological transformation in which AI, automation, robotics and connected systems reshape economies and work.

G

GDPR (General Data Protection Regulation). The European Union's data protection law, which changed the global conversation on data responsibility and made privacy a

board-level concern (Chapter 5).

Generative AI (GenAI). AI systems that create new content, such as text, images, code, audio or video, rather than only analysing existing data. ChatGPT's launch in November 2022 brought GenAI into mainstream public use (Chapter 3).

Ghost users. Employees who rely heavily on AI at work but never disclose it, often fearing their perceived value will diminish. Nearly a third of UK workers surveyed in 2025 admitted to using AI tools secretly (Chapter 4).

Ghost work. Mary L. Gray and Siddharth Suri's term for the hidden human labour that powers AI systems: data labelling, content moderation and microtask QA, typically low-paid, precarious and invisible (Chapter 8).

GPU (Graphics Processing Unit). The high-performance chip class, dominated by Nvidia, on which nearly all advanced AI training and inference runs. Access to GPUs has become a form of strategic currency (Chapter 10).

Green premiums. The additional costs, such as carbon offsets, renewable energy credits and sustainability certification, that larger firms can absorb to claim environmental credentials, potentially squeezing out smaller players (Chapter 12).

Grounding. Connecting a model's outputs to verified sources or constraints so that its answers reflect reality rather than plausible-sounding patterns. Without grounding, broader data access can increase rather than reduce hallucination (Chapter 7).

Guardrails. The technical and procedural constraints placed around an AI system to prevent harmful, false or off-brand outputs.

H

Hallucination. The now-common term for an AI model generating confident but false information: invented legal citations, fabricated research papers, non-existent facts. A product of pattern completion rather than malfunction (Chapters 7 and 9).

Human-in-the-loop (HITL). The design model in which automated systems include human intervention at key stages, during training, operation or review, to ensure control and accountability. The book argues HITL has too often become symbolic: humans present in the process but structurally disempowered (Chapter 9).

Hybrid worker. The book's term for someone whose role evolves through ongoing collaboration with intelligent systems: directing, judging and editing machine output rather than producing everything alone (Chapter 4).

I

Inference. The process, and the cost, of using a trained model to generate an answer. Every prompt triggers real-time computation, and at scale, serving costs can exceed the original training cost (Chapter 12).

Intelligent automation. The combination of AI with process automation so systems

can handle tasks requiring judgment and adaptation, not just repetition. The subject of the original Working Machines (2021).

Interpretability. The research field devoted to decoding what happens inside AI models, using techniques such as activation atlases and vector probing. Increasingly urgent as internal representations grow more abstract: see Neuralese (Chapter 13 and Epilogue).

K

KPI (Key Performance Indicator). A measurable value used to track performance against objectives.

L

Large Language Model (LLM). An AI model trained on vast quantities of text to predict and generate language. LLMs underpin tools such as ChatGPT, Claude, Gemini, Qwen and DeepSeek. They generate what is statistically likely rather than verified truth, which is both their power and their central limitation.

Loadshedding. Scheduled electricity cuts, most prominently in South Africa, used in the book to illustrate how unstable power supply undermines digital and AI access (Chapter 8).

Lock-in (vendor lock-in). Deep dependency on a single provider's models, APIs or cloud ecosystem, making switching expensive or impossible when prices or policies change overnight (Chapter 10).

M

Machine learning. The branch of AI in which systems learn patterns from data rather than following explicitly programmed rules.

Missed generation. The book's term for workers caught in the timing gap of the agentic transition: their roles vanish before new pathways, reskilling routes or redistribution mechanisms exist at scale (Chapter 11).

Model cards. Standardised documentation explaining how a model works: what data it was trained on, what assumptions it carries and what limitations apply. Popularised through Hugging Face's framework and adopted by companies such as Salesforce (Chapter 6).

Model weights. The learned parameters that constitute a trained model. 'Open weights' means these parameters are published, enabling others to run, study and adapt the model (Chapter 10).

Moral outsourcing. The silent transfer of ethical decisions to machines, becoming embedded in how people relate to AI agents at work (Chapter 13).

Multimodal AI. Systems that work across multiple types of input and output, such as text, images, audio and video, rather than a single medium.

MVP (Minimum Viable Product). The simplest version of a product that can be released to test demand. Start-ups have used GenAI to build MVPs, pitch decks and strategies with minimal overhead (Chapter 4).

N

Narrow superintelligence. AI that surpasses human performance in tightly bounded domains, such as legal analysis or market forecasting, without general reasoning beyond that scope (Epilogue).

Natural Language Processing (NLP). The field of AI that enables systems to interpret, analyse and respond to human language, used at executive level to extract insight from unstructured sources such as reviews, filings and transcripts (Chapter 1).

Net zero. Balancing the greenhouse gases emitted with those removed from the atmosphere. The book describes AI's energy appetite as fundamentally in conflict with net-zero commitments (Chapter 12).

Neuralese. The emerging term for the internal, compressed representations AI models develop to process information: efficient for the machine, largely incomprehensible to humans. Neuralese turns transparency into a governance, design and cultural problem, not merely a technical one (Chapter 13).

No-collar worker. The hybrid roles in which human and machine work side by side, each amplifying the other's strengths (Part 3 introduction).

O

OKR (Objectives and Key Results). A goal-setting framework linking objectives to measurable outcomes. Mature data governance shows up in departmental OKRs, not just IT dashboards (Chapter 5).

Open source. Software, and increasingly AI models, whose code or weights are publicly available to use, study and modify. Communities such as Hugging Face act as a counterbalance to centralised model development (Chapters 6 and 10).

Optimisation algorithms. Mathematical tools that find the best solution from many possibilities, used in prescriptive analytics to evaluate courses of action (Chapter 1).

P

Pattern completion. The core mechanism of generative models: predicting what is most likely to come next based on training data. When information is missing, the model fills the gap plausibly rather than pausing, which is the root of hallucination and default overconfidence (Chapter 9).

Predictive modelling. The use of statistical techniques and machine learning to forecast future outcomes from historical data, continuously improving as new data arrives. It shifts leadership from reactive planning to proactive shaping of outcomes (Chapter 1).

Prescriptive analytics. Analytics that go beyond predicting what will happen to recommending what should be done, using optimisation and simulation to propose decisions rather than merely support them (Chapter 1).

Presenteeism. Appearing active and productive at work while leaning heavily on AI tools to do the actual work; reported by 39% of UK workers in 2025 research (Chapter 4).

Prompt. The instruction or input a user gives a generative AI system. The quality and framing of a prompt strongly shape the output.

Prompt engineering. The emerging skill of crafting prompts that reliably produce useful, accurate outputs. Valuable, but no substitute for the judgment needed to evaluate the answer (Chapter 4).

Proof of concept (pilot). A limited trial of a technology before full deployment. On average, organisations abandoned 46% of AI proofs-of-concept before production, a defining pattern of the hype era (Chapter 7).

R

Rare earth elements. The scarce minerals, such as neodymium and dysprosium, essential to AI hardware. China controls about 70% of production, making the ‘virtual’ AI economy dependent on a deeply physical, fragile and geopolitically concentrated supply chain (Chapter 12).

Reality literacy. The predicted core skill of distinguishing real from synthetic and meaningful from mimicked, as essential to the Synthetic Age as media literacy was to the internet era (Epilogue).

Recommender system (recommendation engine). A system that suggests products, content or actions based on user behaviour and preferences. When it misfires, the fix requires cross-functional collaboration, not just data science (Chapter 5).

Red-teaming. Adversarial testing that probes AI systems for potential misuse, bias or failure before and after launch. Now a regular part of responsible development cycles (Chapter 6).

Reinforcement Learning from Human Feedback (RLHF). A training technique in which human preferences guide a model towards safer, more helpful behaviour. Originally a performance technique, later adapted for alignment (Chapter 6).

Resenteeism. Remaining in a role you have grown to resent, in this context because of AI-related changes to the job; reported by 48% of UK workers in 2025 research (Chapter 4).

Reskilling. Training workers for substantially different roles as their current work is automated. The book argues reskilling proved far more fragmented, costly and inaccessible than anticipated, and too often taught tools rather than thinking (Chapter 8).

Responsible AI. The umbrella term for developing and deploying AI with fairness, transparency, safety and accountability built in. Credibility comes from structure, not

statements (Chapter 6).

ROI (Return on Investment). The financial return generated relative to cost. Only 25% of AI initiatives met or exceeded their ROI targets in a 2024 IBM study (Chapter 7).

RPA (Robotic Process Automation). Software that automates repetitive, rule-based digital tasks such as data entry and invoice processing. The bridge between traditional automation and today's AI.

S

Sandboxing. Isolating a system in a controlled environment so its actions, such as a model drafting emails during safety testing, cannot affect the real world (Chapter 7).

Scope 3 emissions. The indirect carbon costs embedded in a company's supply chain. The book predicts businesses will need to map AI-related mineral and water exposure in the same way (Chapter 12).

Sentiment analysis. Using AI to gauge the emotional tone of text, such as customer feedback or social media.

SHAP (SHapley Additive exPlanations) values. A method for interpreting model predictions by showing which inputs had the greatest influence on an output. One of the tools that moved from academic novelty to standard deployment documentation (Chapter 6).

Singularity. The popular-culture notion of AI acting autonomously in its own interest beyond human control. The book treats it as distant, while noting that emergent behaviours in frontier models already warrant pause (Chapter 7).

Skills mismatch. The paradox of millions of jobs being displaced while millions more go unfilled because too few people have the right training: a failure of education and workforce-planning systems to keep pace with technology (Chapter 2).

Sovereign AI. National strategies to build domestically controlled AI capability, data infrastructure and governance, pursued from the Gulf states to the EU's digital sovereignty initiatives (Chapter 10).

Structured and unstructured data. Structured data is organised in defined formats, such as financials and logistics records; unstructured data includes free-form content such as customer feedback, social media, regulations and news. Executive AI systems ingest both (Chapter 1).

Supervised learning. Training a model on labelled examples so it learns the relationship between inputs and correct outputs. The approach that makes human annotation indispensable (Chapter 9).

Synthetic colleague (synthetic identity). An AI agent designed with a consistent voice, personality, backstory and relational presence, embedded in workplaces as a mentor, translator, strategist or communicator. Not pretending to be human, but no longer just a bot (Chapter 13).

Synthetic data. Artificially generated data used to train or test AI systems where real data is scarce, sensitive or biased.

Synthetic media. AI-generated content, including text, images, voice and video, that is increasingly indistinguishable from human-made material (Chapter 13).

Synthetic workspace. An adaptive, AI-mediated virtual environment that responds to a worker's mood, cognitive load and task in real time: the logical endpoint of workplace personalisation (Chapter 13).

T

Token. A small chunk of text, roughly three-quarters of a word on average, that a language model processes. Providers charge by tokens, and 1,000 tokens equate to about 750 words; a single deep working session can consume tens of thousands (Chapter 12).

Training. The process of teaching a model by exposing it to large volumes of data. Training a frontier model can cost upwards of US\$100 million, before the ongoing cost of inference (Chapter 12).

Triage (AI triage). Using AI to prioritise cases, such as patients, tickets or claims, by urgency. Hospital triage systems have been re-evaluated after evidence of demographic bias (Chapter 7).

U

Uncanny valley of authorship. The book's term for the unsettling moment when an AI system produces something eerily close to what you would have written yourself, leaving you unsure how much of your work is truly yours (Chapter 4).

Universal Basic Income (UBI). A policy of unconditional regular payments to all citizens, increasingly debated as a response to AI-driven displacement (Chapters 8 and 11, Epilogue).

User-in-the-loop (UIL). The book's proposed structural pivot from symbolic internal oversight to a transparent, responsive, user-integrated layer of accountability: feedback that is visible, trackable, auditable and able to influence how AI develops. In UIL thinking, the system exists to serve the human, not the reverse (Chapter 9 and Conclusion).

UX (User Experience). The overall experience a person has when interacting with a product or system. UX writers and designers now sit at the centre of AI deployment, shaping how systems communicate (Chapter 3).

V

Vector search (vector representation). Storing and comparing information as numerical vectors that capture meaning, enabling AI systems to retrieve conceptually similar content. Part of the middleware layer connecting AI applications (Chapter 10).

W

Wisdom of crowds. James Surowiecki's concept that distributed knowledge across many people often surpasses centralised expertise, provided the crowd is diverse and independent. Both the promise behind user feedback and, as Birdwatch showed, a mechanism that fails when inputs are gamed (Chapter 9).

Abbreviations

4IR Fourth Industrial Revolution

AI Artificial Intelligence

API Application Programming Interface

AWS Amazon Web Services

CCPA California Consumer Privacy Act

CDO / CDAO Chief Data Officer / Chief Data and Analytics Officer

CEO / CFO / CIO / CTO Chief Executive / Financial / Information / Technology Officer

CRM Customer Relationship Management

CSR Corporate Social Responsibility

DRC Democratic Republic of the Congo

ESG Environmental, Social, and Governance

EU European Union

EUV Extreme Ultraviolet (lithography)

FOMO Fear of Missing Out

GBV Gender-Based Violence

GDPR General Data Protection Regulation

GenAI Generative Artificial Intelligence

GPS Global Positioning System

GPU Graphics Processing Unit

HITL Human-in-the-Loop

HR Human Resources

IEA International Energy Agency

ILO International Labour Organisation

IMDA Infocomm Media Development Authority (Singapore)

IT Information Technology

KPI Key Performance Indicator

LLM Large Language Model

LLP Limited Liability Partnership

MVP Minimum Viable Product

NIST National Institute of Standards and Technology (US)

NLP Natural Language Processing

NTSB National Transportation Safety Board (US)

OBBBA One Big Beautiful Bill Act (US)

OECD Organisation for Economic Co-operation and Development

OKR Objectives and Key Results

OLPC One Laptop Per Child

ORA Office of Responsible AI (Microsoft)
PR Public Relations
QA Quality Assurance
R&D Research and Development
RLHF Reinforcement Learning from Human Feedback
ROI Return on Investment
RPA Robotic Process Automation
SHAP SHapley Additive exPlanations
SQL Structured Query Language
TSMC Taiwan Semiconductor Manufacturing Company
UAE United Arab Emirates
UBI Universal Basic Income
UIL User-in-the-Loop
UK / US United Kingdom / United States
UX User Experience
VC Venture Capital
WEF World Economic Forum
WHO World Health Organization