What is agentic Artificial Intelligence (AI) ? Explain its working. Describe its applications with suitable examples. Discuss the advantages, risks and challenges associated with agentic AI systems.



MAINS MASTERY 2026 · GS PAPER III · QUESTION 16 · STUDENT EDITION

Agentic Artificial Intelligence: Autonomy and Accountability

THE QUESTION, AS SET IN UPSC CSE MAINS 2026

What is agentic Artificial Intelligence (AI) ? Explain its working. Describe its applications with suitable examples. Discuss the advantages, risks and challenges associated with agentic AI systems.

Directive: Explain · Describe · Discuss | 15 marks | 250 words

Science and technology: Agentic AI

1 **PYQ Deep Decode**
Demand, agency and human control

2 **PYQ Intelligence**
AI developments and related PYQs

3 **Additional information**
Working, examples and safeguards

4 **Pitfalls and variations**
Six traps and changed directives

5 **The model answer**
251-word four-part answer

6 **Other ways to write it**
Alternative answer structures

7 **Quick recall**
Sixteen facts for rapid revision

8 **Write it yourself**
Checklist, sources and practice

Guranchal Ma'am, Founder, UPSC Gurus · upscgurus.in

What is agentic Artificial Intelligence (AI) ? Explain its working. Describe its applications with suitable examples. Discuss the advantages, risks and challenges associated with agentic AI systems.

1 PYQ Deep Decode

Read the question the way the examiner set it: the demand, the hidden demand, the structure that follows from both, and how the answer should be presented, opened and closed.

1.1 The demand

The question asks **four things in sequence**: what agentic AI is; how it works; its applications with suitable examples; and its advantages, risks and challenges. **"Explain"** requires making its autonomous working clear, **"Describe"** requires concrete applications, and **"Discuss"** requires a balanced assessment of its benefits and the problems of its deployment. The answer must distinguish **AI that generates a response** from **AI that can pursue a goal, use tools and take actions**. The underlying question is not merely what these systems can accomplish, but what happens when their decisions have real-world consequences.

Words in the question	What they ask you to do	What goes wrong if you miss them
What is agentic Artificial Intelligence (AI)?	Define agentic AI and identify what makes it distinctive: goal-directed behaviour, relative autonomy, planning and the capacity to act through tools.	Defining all AI or generative AI without explaining agency.
Explain its working	Show the sequence from a goal to information gathering, planning, tool use, action and feedback. Explain how the agent can adjust its course.	Writing only that AI uses algorithms, neural networks or machine learning.
Describe its applications	Identify distinct areas of use: administration, business, scientific research, software development and personal services.	Providing a generic list of AI applications without showing autonomous tasks.
with suitable examples	Attach specific, recognisable examples to applications. Distinguish an actual product or documented initiative from a proposed use case.	Claiming that ordinary chatbots, medical image classifiers or predictive models are necessarily agentic.
Discuss the advantages	Explain what autonomy adds: multi-step automation, efficiency, adaptability, continuous operation and potential productivity gains.	Making unsupported claims that agents are invariably faster, cheaper or more accurate than people.
risks and challenges	Assess both direct harms and deployment obstacles: security, privacy, reliability, bias, accountability, interoperability, infrastructure, cost and skills.	Treating cybersecurity as the only problem, or presenting technical limitations without their consequences.
associated with agentic AI systems	Keep every assessment specific to systems that can perform actions, interact with external services and sometimes operate with limited supervision.	Writing an essay about AI in general, including unrelated debates without connecting them to autonomous action.

THE DEMAND IN ONE LINE

Define agentic AI, explain how a goal becomes autonomous action through planning, tools and feedback, demonstrate applications with concrete examples, and weigh productivity against the safety, reliability and governance problems created by that autonomy.

What is agentic Artificial Intelligence (AI) ? Explain its working. Describe its applications with suitable examples. Discuss the advantages, risks and challenges associated with agentic AI systems.

1.2 The hidden demand

When AI systems can independently take consequential actions, where should the boundary of human control lie, and who remains accountable for their decisions?
This is the central governance issue beneath the technical question. A system that drafts a reply presents one category of risk; a system authorised to **send that reply, transfer money, alter a database or execute code** presents another. Its ability to act changes the consequences of error, manipulation and inappropriate decisions. The answer therefore needs a small final part on **human oversight, bounded authority and accountability**.
The distinction is **not between complete automation and complete human control**. Routine, reversible actions may be executed within predefined limits; high-impact, irreversible or rights-affecting decisions require stronger authorisation and review. Responsibility remains with the people and institutions that design, deploy, authorise and operate the system. A system's autonomy does not itself make it a legally accountable person.

A useful test is to ask what the agent is allowed to do, what could happen if it is wrong, and who can intervene or provide a remedy.

Agentic action	Boundary of autonomy	Accountability principle
Classify and route grievances	May categorise cases and prepare replies; disputed disposal and rights-affecting decisions require authorised review.	The public authority remains answerable for the decision and grievance remedy.
Manage customer refunds	May verify information and prepare a refund; payment limits, exception handling and appropriate approvals constrain execution.	The service provider remains responsible for customer protection and transactions.
Run software maintenance	May test and propose code changes; production deployment needs controlled permissions, testing and approval appropriate to risk.	The deploying organisation owns operational security and recovery.
Support clinical decisions	May collect evidence and recommend follow-up; consequential diagnosis or treatment decisions remain subject to qualified clinical judgment.	Professional, organisational and applicable legal responsibilities continue.

WEAK
Agentic AI performs tasks without humans. It improves efficiency but can make mistakes, so governments should regulate it.

STRONG
An AI agent may independently retrieve records and prepare a welfare-benefit decision, but an adverse eligibility determination should remain subject to authorised human review, recorded reasons and appeal. Autonomy can improve processing without transferring public accountability to software.

1.3 The structure

The question has **four demands in a fixed order**. The first three are explicit: meaning and working; applications; advantages, risks and challenges. The fourth is the **small hidden demand**, human oversight and accountability. The shares below guide the body between the introduction and conclusion; they are not a UPSC-prescribed allocation.

What is agentic Artificial Intelligence (AI) ? Explain its working. Describe its applications with suitable examples. Discuss the advantages, risks and challenges associated with agentic AI systems.

Introduction	1. Meaning and working about three-tenths	2. Applications and examples about three-tenths	3. Advantages, risks, challenges about three-tenths	4. Human oversight about one-tenth	Conclusion

Part 1 Meaning and working of agentic Artificial Intelligence — about three-tenths

Meaning and defining characteristics	Goal-directed AI with relative autonomy, context awareness, planning and tool-enabled action; distinguish it from response-only AI.
Goal-directed planning	Interpret a human-set objective; break it into subtasks; select and revise a sequence of steps.
Tools, actions and feedback	Retrieve data, call application programming interfaces (APIs), operate interfaces, evaluate results and replan.
Degree of autonomy	Operate within assigned permissions and boundaries; distinguish supervised from more autonomous execution.

Part 2 Applications of agentic AI with suitable examples — about three-tenths

Administrative and business applications	Grievance triage, customer support, supply-chain coordination and enterprise workflows; Microsoft Copilot Studio agents and Indian enterprise-agent initiatives.
Scientific and technical applications	Software development, automated testing, algorithm discovery and optimisation; Claude Code and Google DeepMind's AlphaEvolve.
Personal and service applications	Research, web navigation, form filling and scheduling; Operator and ChatGPT agent as documented examples.
Suitable examples	Explain what each agent can do, not merely name a product. Label hypothetical applications as illustrative.

Part 3 Advantages, risks and challenges associated with agentic AI systems — about three-tenths

Advantages and productivity gains	Reduced repetitive coordination, multi-step processing, continuous operation and better use of specialist time.
Safety and security risks	Prompt injection, malicious tool use, excessive permissions, data leakage and cascading failures.
Reliability and accountability	Hallucinations, incorrect plans, uncertain outputs, opaque decisions and disputed responsibility.
Governance and implementation challenges	Testing, standards, infrastructure, interoperability, skills, privacy compliance and employment transition.

What is agentic Artificial Intelligence (AI) ? Explain its working. Describe its applications with suitable examples. Discuss the advantages, risks and challenges associated with agentic AI systems.

Part 4 **Human oversight and accountability for autonomous AI actions** **about one-tenth, hidden demand**

Authorisation	Least-privilege access, scope restrictions and spending or action limits.
Human oversight	Approval gates for consequential actions; monitoring and intervention for lower-risk tasks.
Accountability	Audit trails, named institutional responsibility, contestability and remedies.

1.4 How to present it

The form should mirror the **four-part demand**. A brief definition and **goal-to-action flow** explain how agentic AI works. **Headed points with named examples** establish its applications. A compact **benefit-risk-challenge assessment** provides balance. The final oversight point settles the problem of consequential autonomous actions. For a 250-word answer, an elaborate architecture diagram or a large table would use space better spent on examples and analysis.

Form	Fit here	Reason
Definition, short flow, headed points and paired assessment	**Best**	Directly answers all four parts in order; makes working visible and permits specific examples and evaluation.
Four-part mini-table	Good	Each part can be a table row, but technical explanation and examples may become cramped.
Long technical architecture diagram	Limited	Useful for a question specifically on architecture, but too expensive in words and space here.
General essay on AI's future	Weak	Usually misses autonomous working, concrete agentic examples and responsibility for actions.

1.5 The introduction and the conclusion

The introduction serves the whole question. Because the examiner asks what agentic AI is, begin with a compact definition that conveys both autonomy and action. The definition should prepare the reader for working, applications and consequences, rather than describing machine learning generally.

The conclusion settles the whole demand. Agentic AI's usefulness depends on combining goal-directed automation with reliability, security, bounded permissions and accountable human oversight. The concluding judgment should balance innovation and control, at approximately the length of the introduction.

What is agentic Artificial Intelligence (AI) ? Explain its working. Describe its applications with suitable examples. Discuss the advantages, risks and challenges associated with agentic AI systems.

2 PYQ Intelligence 2013 to 2026

The development of agentic AI before the August 2026 examination, the skills UPSC has repeatedly tested in questions on emerging technologies and digital security, and what to prepare next.

2.1 Why the subject was topical in 2026

Between 2024 and August 2026, developments moved AI beyond generating answers towards **planning, using tools and executing tasks**. Commercial agents appeared in software development, research, web navigation and enterprise operations. India advanced its AI infrastructure and governance agenda, while NITI Aayog identified agentic AI as an opportunity for technology services. Security researchers increasingly examined the distinctive risks of autonomous agents. These developments provide **context, not evidence of why UPSC selected the question**. UPSC does not publicly establish such causation. Every event below predates the August 2026 GS examination.

When	What happened	Why it matters for this question
7 March 2024	The Union Cabinet approved the **IndiaAI Mission** with an outlay of **₹10,371.92 crore**, covering computing infrastructure, indigenous AI development, datasets, applications, skilling, startup financing and Safe & Trusted AI (Cabinet/PIB).	India's enabling ecosystem: computation, data, skills and responsible deployment. The Mission is for AI broadly, not exclusively agentic AI.
21 October to 19 November 2024	Microsoft announced autonomous-agent capabilities for **Copilot Studio** on 21 October, followed by public preview availability in November. Agents could respond to events and orchestrate workflows using connected enterprise systems. Microsoft also described business trials, including Pets at Home.	Business applications move beyond answering questions to coordinating processes and taking actions.
23 January 2025	OpenAI introduced **Operator** as a research preview of a computer-using agent able to navigate websites through visual interaction, clicking, typing and scrolling. The accompanying system card discussed model mistakes, prompt injection and user oversight.	A concrete example of the perception, planning, action and safety problems of browser agents.
24 February 2025	Anthropic introduced **Claude Code** in limited research preview, allowing developers to delegate substantial engineering tasks through a terminal interface.	Software development demonstrates autonomous task execution, testing and developer-supervised workflows.
14 May 2025	Google DeepMind announced **AlphaEvolve**, a Gemini-powered evolutionary coding agent combining language-model proposals with automated evaluators for algorithm discovery and optimisation.	Scientific applications need not be mere prediction: agents can propose candidates, test them and iterate.
17 July 2025	OpenAI introduced **ChatGPT agent**, integrating Operator's computer-use functionality with research and tool-based task execution.	Personal and professional assistants increasingly combined research with action, subject to user control.

What is agentic Artificial Intelligence (AI) ? Explain its working. Describe its applications with suitable examples. Discuss the advantages, risks and challenges associated with agentic AI systems.

When	What happened	Why it matters for this question
5 November 2025	**MeitY released the India AI Governance Guidelines** under the IndiaAI Mission. The framework contains **seven guiding Sutras**, recommendations across **six governance pillars**, an action plan and guidance for developers, industry and regulators.	India's broader governance response: human-centricity, risk management, transparency and institutional responsibility.
13 to 14 November 2025	The **Digital Personal Data Protection Rules, 2025**, bear the notification date 13 November (G.S.R. 846(E)) and were published/notified in the Gazette on 14 November, with staggered commencement. They operationalise the DPDP Act, 2023, in phases.	Autonomous data access and processing must respect applicable legal obligations; notification does not mean every requirement applied immediately.
9 December 2025	The OWASP GenAI Security Project published the **Top 10 for Agentic Applications 2026**, identifying risks specific to systems that autonomously plan, use tools and interact with other systems.	Agent security extends beyond hallucinations to goal hijacking, tool misuse, privilege abuse and poisoned context.
11 February 2026	NITI Aayog's Frontier Tech Hub published **India's Technology Services: Reimagination Ahead**. Its **Agentic AI play** proposed productised services, agent-led solutions, legacy modernisation and hybrid human-agent operating models.	Direct Indian policy and industry relevance: productivity, economic opportunity, workforce transition and implementation enablers.
16 to 21 February 2026	India hosted the **India AI Impact Summit 2026** at Bharat Mandapam, New Delhi. The Government's subsequent March report records the event as running from 16 to 21 February, with a focus on the developmental and governance implications of AI.	Demonstrates the importance of responsible, inclusive and development-oriented AI deployment, not a separate legal regime for agents.
17 February 2026	The US National Institute of Standards and Technology (NIST) announced its **AI Agent Standards Initiative**, focusing on secure agent operations, interoperability, authentication and standards.	The distinctive challenge of agents interacting with external tools and systems becomes an explicit standardisation issue.
25 March 2026	MeitY informed Parliament that **10 Indian startups** were selected for the IndiaAI Startups Global programme. The list included **Voicing AI**, described as an enterprise-grade intelligent voice-agent platform.	An Indian agent-oriented enterprise example. Selection for a programme is not proof of deployment at scale or independently measured outcomes.
20 August 2026	The **Technology Development Board, Department of Science and Technology**, announced an agreement to support commercialisation of **Fixit**, an indigenous agentic AI platform developed by Delhi-based One2X Tech, intended for coordinated enterprise workflows.	A directly relevant Indian initiative shortly before the examination. Commercialisation support must not be misrepresented as completed, nationwide deployment.

Two distinctions matter. First, the IndiaAI Mission and the AI Impact Summit concern AI generally, although they help create the environment in which agentic systems develop. Second, commercial announcements, research demonstrations, startup selections and government-supported commercialisation are different evidentiary categories. They should not be presented as equivalent proof of successful real-world adoption.

What is agentic Artificial Intelligence (AI) ? Explain its working. Describe its applications with suitable examples. Discuss the advantages, risks and challenges associated with agentic AI systems.

2.2 The skill UPSC keeps testing

UPSC has repeatedly combined **technological understanding with applications, benefits and vulnerabilities**. In 2013 it asked how an emerging technology worked and what its advantages and disadvantages were. Subsequent questions examined cyber threats, data protection and responsible use. In 2023, the paper explicitly connected artificial intelligence with clinical applications and privacy. In 2026, the demand moved to **AI capable of performing actions**, making the boundary between technological capability and human accountability especially important. The questions below are reproduced verbatim, including the original punctuation and spacing.

Year	The question as UPSC set it	What it asked of you	What changed in 2026
2013 GS III, Q16(b)	How does the 3D printing technology work ? List out the advantages and disadvantages of the technology.	Explain an emerging technology's operation, strengths and limitations.	The same explanatory discipline now applies to autonomous AI, with applications and governance added.
2017 GS III, Q9	Discuss the potential threats of Cyber attack and the security framework to prevent it.	Connect digital threats with the safeguards needed to prevent them.	Agents introduce new attack routes through tool access, permissions and automated actions.
2018 GS III, Q19	Data security has assumed significant importance in the digitized world due to rising cyber crimes. The Justice B. N. Srikrishna Committee Report addresses issues related to data security. What, in your view, are the strengths and weaknesses of the Report relating to protection of personal data in cyber space?	Evaluate privacy safeguards rather than merely describe a technology.	Agentic systems may access and combine information across multiple applications, sharpening privacy concerns.
2021 GS III, Q10	Keeping in view India's internal security, analyse the impact of cross-border cyber attacks. Also discuss defensive measures against these sophisticated attacks.	Assess cyber threats and matched defensive measures.	Autonomous agents can expand the consequences of compromised credentials and malicious instructions.
2022 GS III, Q19	What are the different elements of cyber security ? Keeping in view the challenges in cyber security, examine the extent to which India has successfully developed a comprehensive National Cyber Security Strategy.	Identify security elements, institutional preparedness and gaps.	Securing agents requires controls over their identities, tool access, behaviour and monitoring.
2023 GS III, Q5	Introduce the concept of Artificial Intelligence (AI). How does AI help clinical diagnosis? Do you perceive any threat to privacy of the individual in the use of AI in healthcare?	Define AI, demonstrate an application and evaluate its privacy implications.	The new question requires explaining the additional capacity to plan and act, with wider applications and risks.
2024 GS III, Q10	Describe the context and salient features of the Digital Personal Data Protection Act, 2023.	Understand the statutory framework for personal data protection.	Privacy and lawful processing become implementation constraints on agents using personal data.
2026 GS III, Q9	Explain how fake news and disinformation pose threat to Internal Security and Public Order in Indian context? In this regard, discuss salient features of amendments in respect of Information Technology (Intermediatory Guidelines and Digital Media Ethics Code) Rules 2021.	Explain digital harms and examine a regulatory response.	The same paper tests the governance of digital risks alongside the consequences of autonomous AI.

What is agentic Artificial Intelligence (AI) ? Explain its working. Describe its applications with suitable examples. Discuss the advantages, risks and challenges associated with agentic AI systems.

Also related, verbatim: 2013 GS III, Q22: "What are social networking sites and what security implications do these sites present ?" 2023 GS III, Q2: "What is the status of digitalization in the Indian economy? Examine the problems faced in this regard and suggest improvements."

CARRY THIS METHOD TO ANY EMERGING-TECHNOLOGY QUESTION

1. **Define** the technology through the capability that distinguishes it from earlier systems.
2. **Explain** its working as a sequence of understandable operations, not as a collection of technical jargon.
3. **Demonstrate** its uses through examples that show what the technology actually changes.
4. **Assess** benefits alongside the risks and constraints created by that distinctive capability.
5. **Conclude** with the conditions for responsible adoption and the institutions responsible for meeting them.

2.3 What to prepare next

TECHNOLOGIES, INSTITUTIONS AND LEGAL INSTRUMENTS

IndiaAI Mission, 2024	Cabinet approval, 7 March 2024; ₹10,371.92 crore outlay; seven programme pillars, including Safe & Trusted AI.
India AI Governance Guidelines, 2025	MeitY, 5 November 2025; seven Sutras, six governance pillars, proportionate risk management and human-centricity. A policy framework, not a standalone agentic AI statute.
DPDP Act, 2023 and Rules, 2025	Personal-data governance; Rules dated 13 November 2025, Gazette publication 14 November, with phased commencement. Apply obligations according to their operative dates.
NITI Aayog roadmap, February 2026	Agent-led services, enterprise modernisation, human-agent models, infrastructure, talent, R&D and policy enablers.
OWASP Agentic Top 10, December 2025	Security-risk guidance for autonomous AI applications; not legislation or a mandatory Indian standard.
NIST initiative, February 2026	Agent identity, authentication, secure operation and interoperability; standards development, not certification of all agents.

CONCEPTS AND DISTINCTIONS

Traditional AI versus generative AI versus agentic AI	Recognition or prediction; content generation; goal-directed, tool-enabled action. These categories can overlap.
Workflow automation versus agency	A fixed script follows prescribed steps; an agent can select and revise steps within its authorised task.
Reasoning versus acting	A plan is not a completed action. External tool calls distinguish many practical agents.
Prompt injection	Untrusted content attempts to redirect an agent from authorised instructions.
Human-in-the-loop and human-on-the-loop	Approval at critical actions versus supervision of ongoing bounded activity.
Least privilege and auditability	Give agents only necessary permissions; retain usable records of their actions and outcomes.

What is agentic Artificial Intelligence (AI) ? Explain its working. Describe its applications with suitable examples. Discuss the advantages, risks and challenges associated with agentic AI systems.

3 Additional information

Everything needed to answer this question well. Tier 1 contains the definition, working, applications and balanced assessment directly required by the question. Tier 2 provides the wider technical, economic, ethical and governance understanding needed to distinguish agentic AI from ordinary automation.

TIER 1 · ESSENTIAL: MEANING, WORKING, APPLICATIONS, ADVANTAGES AND RISKS

What makes an AI system agentic?

Agentic artificial intelligence refers to AI systems designed to pursue assigned objectives with a degree of autonomy by interpreting their environment, planning steps, using available tools, taking actions and responding to feedback. The objective is ordinarily supplied or authorised by a person or organisation; the agent may choose intermediate subgoals and actions within that mandate. **Autonomy is relative, not necessarily unrestricted.**

Characteristic	What it means	Illustration and distinction
Goal-directed behaviour	The system works towards an objective, rather than generating only an immediate response to a question.	Instead of merely describing travel options, an authorised agent can search alternatives and prepare an itinerary.
Planning and decomposition	It breaks a complex task into smaller steps and determines an execution sequence.	For software repair: inspect the repository, identify the defect, write a patch, test it and report the result.
Perception and context	It reads relevant documents, application states, tool outputs or other observations to understand the task's current situation.	A browser agent reads the updated contents of a webpage before deciding where to click next.
Tool use and action	It can call APIs, search databases, operate a browser or invoke authorised software functions.	Submitting an approved support-ticket update is an action; generating suggested wording alone is not.
Feedback and replanning	It checks outcomes and may change its approach when a step fails or new information appears.	If a code test fails, the agent reviews the error and proposes a revised patch.
Memory or maintained state	It may retain the task's intermediate results, previous steps or permitted persistent context.	A service agent can avoid repeating completed verification steps. Persistent memory is useful but not essential to every agent.
Variable autonomy	Permissions and supervision determine whether it recommends, acts after approval or completes bounded tasks independently.	An agent may draft a refund autonomously but require approval to disburse the money.
Use of AI models	Language models can interpret instructions, reason over context and select tools; other components can perform retrieval, checks, deterministic calculations or execution.	An agent is usually a **system built around models and tools**, not merely a language model with a new label.

How an agent works: from an objective to a verified result

1. Goal Human-set objective and permissions → **2. Observe** Read context, data and environment → **3. Plan** Break into tasks and select steps → **4. Act** Use tools within authorised limits → **5. Check** Evaluate results, revise or escalate

What is agentic Artificial Intelligence (AI) ? Explain its working. Describe its applications with suitable examples. Discuss the advantages, risks and challenges associated with agentic AI systems.

A teaching flow, not a single mandatory software architecture. The fifth stage feeds back into observation and planning until the objective is met, an execution limit is reached or the task is stopped. A consequential action may require human authorisation before execution. The flow is consistent with documented computer-using agents, coding agents and agent-security guidance.

Stage	What happens technically	Running example: an authorised customer-service agent
1. Receive the objective	A user, event or authorised application supplies the task, scope and relevant constraints. The system identifies what successful completion would mean.	"Investigate this customer's delayed order and prepare the appropriate response."
2. Observe the environment	The agent accesses permitted records, tool descriptions and current state; retrieved material is evidence, not automatically a trusted instruction.	Reads order status, delivery tracking and company refund policy.
3. Plan and decompose	An AI model or planning mechanism selects intermediate tasks and decides which tools may help complete them.	Verify the order, establish the cause of delay, check remedies, prepare the response.
4. Execute through tools	The agent calls APIs or other software interfaces; deterministic systems carry out transactions or return requested results.	Queries tracking records and prepares an update to the customer-support system.
5. Evaluate the outcome	Tool responses and validation checks indicate whether the intended action succeeded and whether further work is necessary.	Confirms that the order number matches and that the proposed remedy follows the policy.
6. Replan or escalate	Errors, contradictions or permission limits trigger revised steps, human review or termination.	An exceptional refund goes to an authorised officer rather than being issued automatically.
7. Record and finish	The system records relevant decisions, tool calls, approvals, outputs and status for oversight and recovery.	Stores the case outcome and identifies whether a human approved a consequential action.

Three distinctions that prevent conceptual errors

System	Principal capability	What changes with agentic AI?
Traditional or predictive AI	Classifies, predicts or detects patterns, such as fraud probabilities or diagnostic features.	An agent can combine prediction with authorised decisions and follow-up actions.
Generative AI	Produces text, images, code or other content in response to an input.	An agent can use generation as part of a longer task, call tools and act on results.
Rule-based process automation	Executes predefined conditional steps in a predictable workflow.	An agent can interpret variable conditions and revise steps without every path being explicitly programmed.
Agentic AI	Pursues a goal through observation, planning, tools and feedback, with autonomy determined by system design.	Agency concerns the **organisation and execution of actions**, not proof of consciousness, perfect reasoning or unrestricted intelligence.

What is agentic Artificial Intelligence (AI) ? Explain its working. Describe its applications with suitable examples. Discuss the advantages, risks and challenges associated with agentic AI systems.

Applications: specific examples and the actual agentic capability

Application	Example and verified status	What the agent does that a response-only AI does not
ADMINISTRATIVE AND BUSINESS APPLICATIONS		
Enterprise workflow management	**Microsoft Copilot Studio autonomous agents**: announced October 2024 and placed in preview in November 2024. Microsoft described event-triggered workflows, connected systems and case processing. Its Pets at Home example concerned case preparation for human review.	Responds to an event, retrieves information from systems, coordinates steps and prepares or performs authorised workflow actions.
Indian enterprise agents	**Fixit, One2X Tech**: on 20 August 2026, the Technology Development Board announced support for commercialising an indigenous platform for coordinated enterprise workflows. This is a commercialisation initiative, not proof of fully deployed adoption.	Aims to interpret enterprise objectives and execute coordinated actions through integrations and orchestration.
Voice-based business services	**Voicing AI**: listed by MeitY on 25 March 2026 among ten startups selected under the IndiaAI Startups Global programme, described as an enterprise-grade intelligent voice-agent platform.	Illustrates voice-driven task handling and enterprise-agent services; programme selection alone does not establish measured effectiveness.
Public administration	**Illustrative application:** a district grievance agent could identify the department concerned, gather status records, route the complaint, flag overdue cases and draft updates.	Coordinates multiple case-management steps, while final disposal and rights-affecting determinations remain under authorised public officials.
SCIENTIFIC AND TECHNICAL APPLICATIONS		
Software engineering	**Anthropic's Claude Code**, introduced in research preview on 24 February 2025, enables developers to delegate engineering tasks from a terminal.	Can inspect software, propose edits, invoke permitted development tools and follow through on a multi-step coding task.
Algorithm discovery	**Google DeepMind's AlphaEvolve**, announced 14 May 2025, combines language-model proposals with automated evaluators to discover and optimise algorithms.	Generates candidate solutions, tests them against evaluators and uses feedback to guide subsequent candidates.
Scientific research coordination	**Illustrative application:** a research agent could search literature, select authorised computational tools, run simulations and organise results for a scientist.	Performs a sequence of research operations. Experimental validity and scientific interpretation still require independent assessment.
Cybersecurity operations	**Illustrative application:** an authorised security agent could correlate alerts, retrieve system logs, recommend containment and perform pre-approved low-risk steps.	Connects detection to investigation and bounded response. Unrestricted containment powers could themselves disrupt essential services.
PERSONAL AND SERVICE APPLICATIONS		
Browser-based assistance	**OpenAI's Operator**, introduced on 23 January 2025 as a research preview, used a browser to navigate sites, enter information and complete permitted tasks. Its functionality was subsequently integrated into ChatGPT agent in July 2025.	Interacts with websites rather than only describing what the user should do; consequential steps can require confirmation.

What is agentic Artificial Intelligence (AI) ? Explain its working. Describe its applications with suitable examples. Discuss the advantages, risks and challenges associated with agentic AI systems.

Application	Example and verified status	What the agent does that a response-only AI does not
Personal productivity	**ChatGPT agent**, introduced 17 July 2025, combines research with computer-use and connected-tool capabilities for multi-step online tasks.	Can gather information, work across tools and prepare or execute authorised steps, instead of merely supplying instructions.
Healthcare coordination	**Illustrative application:** an agent could gather reports, organise appointment slots and flag missing follow-up records for clinical review.	Coordinates administrative actions; it should not be presented as an independently authorised doctor or as proof of better diagnoses.
Agriculture and supply chains	**Illustrative application:** an agricultural advisory agent could combine weather alerts, market prices and inventory to prepare crop-marketing or transport options.	Performs multi-source analysis and coordinates follow-up tasks. Sales, payments and safety-critical recommendations need appropriate safeguards.

Use examples precisely. AlphaEvolve is a documented algorithm-discovery agent; Claude Code and Operator are documented agentic tools; Copilot Studio provides enterprise agent capabilities; Fixit received government commercialisation support; Voicing AI was selected for a startup programme. Public-administration, clinical-coordination and agricultural examples above are explicitly **illustrative designs**, not claims of completed government roll-outs.

Advantages, risks and challenges: always connect the mechanism to the consequence

Dimension	Why it arises in agentic AI	Practical significance or response
ADVANTAGES		
Productivity	Agents can execute a chain of related tasks without repeated manual initiation.	May reduce coordination costs and release skilled employees for complex judgment, subject to real-world evaluation.
Continuous operation	Software agents can monitor defined processes and respond to events outside ordinary working hours.	Useful for authorised alerts, workflow routing and service operations; continuous operation still requires safeguards.
Adaptive task completion	An agent can reconsider a plan when the environment changes or a tool reports failure.	Potentially more flexible than a rigid script in variable processes.
Scalability and accessibility	Reusable agent workflows can support many requests and potentially multiple languages or interfaces.	May improve routine service access where language, connectivity, data quality and digital inclusion are addressed.
RISKS		
Prompt injection and goal hijacking	A malicious webpage, document or tool response can contain instructions designed to redirect the agent from its authorised goal.	Treat external content as untrusted data; separate instructions from retrieved material and check proposed actions.
Tool misuse and privilege abuse	An agent with email, payment, database or administrative privileges can perform harmful actions if manipulated or mistaken.	Use least-privilege credentials, scoped APIs, transaction limits and approval gates.
Data leakage and privacy harm	Agents may connect information from several systems and transmit it through external tools.	Control data access, processing purpose, retention, disclosure and cross-system transfers.

What is agentic Artificial Intelligence (AI) ? Explain its working. Describe its applications with suitable examples. Discuss the advantages, risks and challenges associated with agentic AI systems.

Dimension	Why it arises in agentic AI	Practical significance or response
Compounding errors	A mistaken assumption in one step can affect later decisions and automated actions.	Validate intermediate results and stop or escalate where errors become consequential.
Bias and unequal treatment	Data, models and decision rules may reflect systematic exclusions or inaccurate assumptions.	Conduct context-specific fairness tests, especially in welfare, employment, credit and healthcare.
Accountability gaps	Several models, tools, developers and operators may contribute to an action, obscuring why it occurred.	Maintain audit trails, assigned responsibilities and accessible complaint or correction procedures.
IMPLEMENTATION CHALLENGES		
Reliability and evaluation	An agent may succeed on a demonstration while failing under unusual inputs, changing interfaces or incomplete data.	Test whole-task success, harmful actions, recovery and behaviour in realistic operating conditions.
Integration and interoperability	Legacy systems, inconsistent records, fragmented APIs and incompatible authentication restrict coordinated execution.	Build secure integrations, standard interfaces and reliable identity management.
Cost and infrastructure	Repeated model calls, tool use, monitoring and secure execution consume computing capacity and resources.	Evaluate total operating cost, latency, energy use and performance against conventional automation.
Workforce transition	Agents can alter how routine knowledge work, customer support and software services are organised.	Invest in reskilling, job redesign, professional supervision and equitable access to productivity gains.
Standards and governance	Agent identity, permissions, action boundaries and behaviour across organisations remain difficult to standardise.	Use risk-based governance, sectoral oversight and evolving interoperability and security standards.

Safety controls that match specific agentic risks

Control	What it does	Example
Least privilege	Authorises only the tools and records necessary for an approved task.	An appointment agent can check availability but cannot access unrelated financial data.
Human approval gates	Require authorisation before defined consequential, irreversible or high-risk actions.	A proposed transfer is reviewed before execution.
Tool and input isolation	Separates untrusted retrieved material from controlling instructions and restricts unsafe execution.	A malicious instruction embedded in a retrieved PDF is not treated as an order from the user.
Independent validation	Checks essential outputs against policy, deterministic rules or independent evidence.	Software tests must pass before a patch is considered for deployment.
Execution limits	Constrain retries, spending, time, data access or the number of actions.	An agent stops after a defined error threshold and requests review.

What is agentic Artificial Intelligence (AI) ? Explain its working. Describe its applications with suitable examples. Discuss the advantages, risks and challenges associated with agentic AI systems.

Control	What it does	Example
Observability and logs	Preserve tool calls, decisions, approvals and relevant outcomes for investigation.	A public authority can reconstruct why a case was routed or escalated.
Rollback and interruption	Allow operators to stop a process and reverse actions where technically feasible.	A failed inventory update is rolled back; a wrongful external disclosure may not be reversible.
Evaluation and red-teaming	Test agent behaviour under adversarial instructions, unexpected states and realistic workflows.	A security team checks whether a support agent follows malicious instructions inside a customer message.

The benefit is end-to-end task completion. Ordinary generative AI can draft a purchase order. Agentic AI may also inspect approved inventory records, compare suppliers and prepare or submit the order within authorised limits. It is the transition from **content to action** that creates much of its economic value.

The risk is also end-to-end task completion. If the agent misunderstands a supplier record, accepts malicious instructions or exceeds its authority, the mistake may turn into a real transaction. **Authority, validation and accountability must accompany capability.**

TIER 2 · EXTENDED UNDERSTANDING: THE WHOLE TOPIC AROUND IT

Architectural ideas worth understanding

1. **Language model as controller:** interprets goals and context, selects the next action and formulates tool requests. Other kinds of models or planners may also participate.
2. **Tool calling:** structured requests to software functions, search systems, databases or external applications. A model proposing a tool call does not mean the operation was successfully executed.
3. **Retrieval-augmented generation:** supplies relevant information from permitted sources. Retrieval improves access to context but does not guarantee truth or correct decisions.
4. **Task memory:** maintains intermediate state; optional persistent memory can support continuity but creates privacy and manipulation concerns.
5. **Plan-and-execute and reasoning-action patterns:** agents can formulate a plan and then act, or interleave reasoning, observation and action. No single approach is mandatory.
6. **Multi-agent orchestration:** specialist agents may divide tasks, such as research, coding and testing, under a coordinator. More agents do not automatically mean better results.
7. **Deterministic controls:** rules engines, validators and approval systems should constrain decisions that cannot safely depend on unconstrained model output.

Autonomy is a spectrum

1. **Recommendation:** the system analyses a situation and advises a person, who decides whether to act.
2. **Human-in-the-loop:** the system plans and prepares actions, but specified actions require explicit approval.
3. **Human-on-the-loop:** the system performs bounded tasks while an operator monitors, intervenes and can stop it.
4. **More autonomous operation:** the system completes predefined tasks within tightly specified permissions, budgets, controls and escalation rules.
5. **Risk-sensitive boundaries:** a temporary calendar adjustment and a decision to terminate a public benefit are not equivalent; the latter warrants substantially stronger scrutiny.
6. **Stopping conditions:** safe operation requires rules for uncertainty, repeated failure, conflicting instructions and exhausted permissions.

What is agentic Artificial Intelligence (AI) ? Explain its working. Describe its applications with suitable examples. Discuss the advantages, risks and challenges associated with agentic AI systems.

Why agentic systems create new security problems

The **OWASP Top 10 for Agentic Applications 2026**, released in December 2025, reflects the transition from risks associated with generated content to risks associated with **authorised and unauthorised actions**. A malicious instruction can enter through a webpage, email, repository file or tool output and attempt to redirect the agent's objectives. The resulting harm depends on the tools and permissions available to the agent. An attacker need not control the model itself if the system trusts hostile external material.

Threat	Failure pathway	Key defence
Agent goal hijack	Untrusted text redirects the agent away from its approved objective.	Instruction hierarchy, input isolation and policy checks before actions.
Tool misuse	Compromised or mistaken instructions lead to unsafe calls to connected applications.	Tool allowlists, argument validation and restricted operations.
Identity and privilege abuse	Agents inherit overly broad credentials or act beyond their intended authority.	Dedicated identities, least privilege and short-lived credentials where appropriate.
Memory and context poisoning	Malicious or false material persists and influences later steps.	Trusted memory boundaries, provenance controls and review of retained state.
Cascading failure	One incorrect decision affects multiple dependent tasks or agents.	Independent checks, bounded execution, monitoring and recovery procedures.

These are operational threat categories illustrated by OWASP's guidance, not a claim that every agent will experience each failure. The exact safeguards depend on the tools, data, operating environment and potential harm.

How to judge whether an agent actually works

Evaluation question	Useful measure or test
Did it complete the task?	End-to-end success on realistic tasks, not merely fluent explanations or attractive demonstrations.
Was the result correct?	Independent verification of actions and final outputs against specified requirements.
Did it remain authorised?	Rate of prohibited actions, permission violations and unsafe tool calls.
Can it resist manipulation?	Adversarial tests using malicious retrieved documents, tool responses and injected instructions.
How does it handle failure?	Recovery, escalation and safe termination in unfamiliar conditions.
Does it justify its cost?	Human effort saved, operating costs, error correction, latency and required oversight.
Can its conduct be audited?	Quality of logs, action attribution, approval records and reproducibility of relevant decisions.
Is it fair and accessible?	Disaggregated performance, language accessibility, bias tests and effective remedies.

India's policy position: enabling innovation while governing risk

India does not need to be described as having a comprehensive, standalone **Agentic AI Act**. The relevant position before the August 2026 examination is a combination of the **IndiaAI Mission**, existing applicable legislation, sectoral oversight, the **India AI Governance Guidelines (5 November 2025)** and emerging technical safeguards.

What is agentic Artificial Intelligence (AI) ? Explain its working. Describe its applications with suitable examples. Discuss the advantages, risks and challenges associated with agentic AI systems.

Instrument	Its actual role	How it connects to agentic systems
IndiaAI Mission	Cabinet-approved programme of 7 March 2024, with ₹10,371.92 crore outlay and seven pillars.	Builds enabling computing, datasets, research, skilling, applications and responsible-AI capabilities; it is not an agent-specific licensing regime.
India AI Governance Guidelines	MeitY framework released 5 November 2025, grounded in seven Sutras and six governance pillars, with risk-based, human-centric oversight.	Provides policy direction for testing, transparency, accountability and proportionate governance. It is guidance, not a newly enacted comprehensive AI statute.
DPDP Act and Rules	The DPDP Act, 2023, and Rules, 2025, regulate digital personal-data processing within their applicable scope and phased commencement.	Agents using personal information must operate within applicable lawful-processing, security and other obligations; the rules are not an exclusive agentic-AI code.
NITI Aayog technology-services roadmap	February 2026 strategic report identifying agentic AI and hybrid human-agent models as opportunities for the services industry.	Provides an economic and industrial strategy perspective, including skills, infrastructure and innovation; its future projections are scenarios, not observed outcomes.
IndiaAI startup support	Programme selection of Indian AI startups, including Voicing AI, reported to Parliament on 25 March 2026.	Shows ecosystem-building and product development without establishing universal reliability or commercial deployment.
TDB support for Fixit	Commercialisation agreement announced on 20 August 2026 by the Department of Science and Technology's Technology Development Board.	Shows targeted support for an indigenous agentic platform, with enterprise integration, security and scalability among development priorities.

NITI Aayog's opportunity and its limitations

In **India's Technology Services: Reimagination Ahead** (February 2026), NITI Aayog's Frontier Tech Hub identified an **Agentic AI play** for India's technology-services industry. It proposed productised solutions for business functions, agent-led modernisation of legacy systems, solutions for medium-sized enterprises and **hybrid human-agent workforces**. The report highlighted four broad enablers: **infrastructure, talent, R&D and policy**.

This is especially relevant to India because the opportunity is not only to sell more AI software but to reconsider how technology services are delivered. At the same time, **potential future market value is not realised revenue**, and labour substitution is not synonymous with improved welfare. Successful adoption depends on effective systems, trustworthy outcomes, responsible organisational change and large-scale reskilling.

Economic and institutional opportunities

1. **Services transformation:** shift selected tasks from manual, repetitive execution towards outcomes supported by software and human oversight.
2. **Indian-language access:** multilingual agents may lower barriers to routine service interaction where reliable language capabilities and accessibility are provided.
3. **MSME productivity:** affordable, appropriately scoped agents may assist smaller enterprises with inventory, support and documentation.

Conditions that could limit these gains

1. **Computing and infrastructure:** repeated inference, secure execution and tool use impose operational costs.
2. **Data quality:** incomplete, conflicting or inaccessible records can undermine even capable models.
3. **Digital exclusion:** language, connectivity, disability access and unequal technical capacity can limit benefits.
4. **Workforce disruption:** changing demand for routine knowledge work requires transition support, not automatic assumptions about net job creation.

What is agentic Artificial Intelligence (AI) ? Explain its working. Describe its applications with suitable examples. Discuss the advantages, risks and challenges associated with agentic AI systems.

4. **Public service support:** agent-assisted routing, information retrieval and case preparation may improve responsiveness without delegating statutory discretion.
5. **Research and innovation:** coding and algorithm-discovery agents can support experimentation and technical development.
5. **Vendor and platform dependence:** concentration of computing, models and proprietary integrations can restrict choice and interoperability.
6. **Liability and redress:** affected individuals need identifiable institutions capable of correcting mistakes and providing remedies.

Two debates worth knowing

Should high-impact decisions ever be fully autonomous?

One view emphasises speed, scale and consistency: properly tested agents could automate many decisions more efficiently than manual processes. The opposing concern is that opaque, fallible systems can deny benefits, expose data or trigger irreversible consequences without meaningful recourse. The balanced position is **proportionate autonomy**: permit bounded low-risk execution, but require stronger authorisation, independent checks, recorded reasons and contestable decisions where safety or rights are at stake.

Will agentic AI create productivity or undermine employment?

Supporters expect agents to reduce routine work, help enterprises scale and enable new high-value services. Critics emphasise displacement of routine knowledge work, concentration of value among large technology firms and unequal access to computing and skills. The balanced position is a **human-agent transition**: measure actual productivity, redesign work responsibly, support reskilling and distribute benefits rather than treating labour replacement as the sole objective.

4 Pitfalls and variations

The mistakes that cost marks in this question, the claims never to make, and how the answer must change if the examiner changes the directive or narrows the subject.

4.1 Six pitfalls, each with its correction

Confusing agentic AI with generative AI

~~Agentic AI is AI that generates human-like text, images and answers.~~

Define it through goal-directed autonomy, planning, tool-enabled actions and feedback. Generative models may form one component of an agent.

No working mechanism

~~Agentic AI works using machine learning, large datasets and neural networks.~~

Explain the goal, observation, planning, tool use, verification and replanning cycle. This directly answers "Explain its working".

Applications without agentic examples

~~AI is used in healthcare, agriculture, banking and education.~~

Show an autonomous task in each selected area and name documented examples such as Claude Code, AlphaEvolve, Copilot Studio or Operator.

Benefits without assessment

~~Agents always improve accuracy, eliminate errors and replace manual intervention.~~

State conditional productivity and flexibility gains, then assess mistakes, security risks, oversight costs and reliability limitations.

What is agentic Artificial Intelligence (AI) ? Explain its working. Describe its applications with suitable examples. Discuss the advantages, risks and challenges associated with agentic AI systems.

Risks reduced to privacy alone

~~The main challenge of agentic AI is that personal data may be leaked.~~

Add prompt injection, tool misuse, privilege abuse, cascading errors, bias, unreliable execution and accountability gaps.

No human responsibility

~~If an autonomous agent makes a decision, the AI itself should be held responsible.~~

Distinguish technical autonomy from legal and institutional accountability; specify scoped permissions, approval gates, logs and remedies.

4.2 Never write these

- "Agentic AI and generative AI are the same technology."
  Generative models may be used within agents, but agency concerns goal-directed action, tool use and feedback rather than content generation alone.
- "Agentic AI always acts without human instructions or intervention."
  Objectives, permissions and supervision are usually assigned by people; the extent of autonomy varies.
- "An AI chatbot is automatically an agentic AI system."
  Answering questions or generating text does not by itself establish autonomous planning and action.
- "AlphaFold and AlphaEvolve are interchangeable examples of agentic AI."
  AlphaFold is principally associated with protein-structure prediction; AlphaEvolve, announced in May 2025, is the documented agentic algorithm-discovery example.
- "India has enacted a comprehensive Agentic AI Regulation Act."
  The India AI Governance Guidelines were released in November 2025 as a policy framework; they are not such an Act.
- "The DPDP Rules, 2025, made every data-protection obligation immediately enforceable."
  The Gazette notification provides staggered commencement. The original notification is dated 13 November 2025 and was published on 14 November.
- "The IndiaAI Mission was created exclusively to develop autonomous AI agents."
  It covers the wider AI ecosystem, including computing capacity, datasets, foundational models, applications, talent and trustworthy AI.
- "Government support for Fixit proves agentic AI has already been deployed across India's public administration."
  The Technology Development Board announced commercialisation support on 20 August 2026, not a nationwide administrative rollout.
- "Agentic AI guarantees faster and more accurate results than human employees."
  Performance depends on task complexity, system design, data quality, verification, operational costs and human oversight.
- "Human oversight removes all risks of autonomous AI."
  Oversight can fail or become ineffective; sound governance also requires technical restrictions, testing, monitoring, recourse and recovery.

What is agentic Artificial Intelligence (AI) ? Explain its working. Describe its applications with suitable examples. Discuss the advantages, risks and challenges associated with agentic AI systems.

4.3 If the question changes

If it asks	The structure becomes	What must change
Explain the architecture of agentic AI	Agent components; model and planner; memory; tools; observation-action-feedback cycle; execution safeguards.	Give the technical architecture and information flow most space. Applications become supporting illustrations.
Differentiate agentic and generative AI	Definitions; autonomy, planning, tool use, memory and output distinctions; paired examples; areas of overlap.	Use a comparison table. Avoid claiming that the categories are mutually exclusive.
Critically evaluate agentic AI in public administration	Governance use cases; efficiency; exclusion and due-process risks; privacy; institutional safeguards; judgment.	Focus on public accountability, citizen rights and administrative discretion rather than generic business uses.
Analyse the cybersecurity threats of AI agents	Attack surface; prompt injection; tool misuse; identity abuse; cascading harm; defence-in-depth.	Place OWASP's agentic security framework, NIST standards and least-privilege design at the centre.
Discuss the impact on employment	Task automation; productivity; displacement; skill changes; distributional effects; workforce transition.	Use the NITI Aayog February 2026 roadmap and separate projected opportunities from realised labour outcomes.
Evaluate India's readiness for agentic AI	Compute and data; research and startups; public-sector opportunities; regulation and safety; skills and inclusion.	Use the IndiaAI Mission, governance guidelines, NITI roadmap and dated Indian initiatives. Assess readiness, not just announcements.
Examine ethical and legal accountability	Autonomous action; bias and privacy; responsibility allocation; human review; auditability; redress.	The hidden demand becomes the main body, with rights-affecting decisions and safeguards examined in depth.
The same subject for **10 marks**	Definition; compact working flow; two or three examples; paired benefits and risks; one oversight line.	Compress examples and explanation while preserving every explicit demand.

What is agentic Artificial Intelligence (AI) ? Explain its working. Describe its applications with suitable examples. Discuss the advantages, risks and challenges associated with agentic AI systems.

5

The model answer

One model answer, written to the UPSC Gurus house range of 250 to 253 words (**251 words exactly**, counting every whitespace-separated token in the answer sheet, including section numerals, headings, point labels and flow-box text; UPSC's printed limit is 250 words). The answer follows the question's four demands in order: meaning and working, applications with examples, advantages and risks, and a short final part on human oversight. A compact flow makes autonomous working visible; headed points connect applications to actual examples and distinguish productivity gains from risks and implementation challenges. The introduction defines agentic AI; the conclusion reconciles autonomy with accountability.

Model answer ## A working flow, applications and balanced assessment

251 words

Agentic artificial intelligence comprises goal-directed systems that plan and execute authorised actions using tools and feedback. Unlike response-only generative AI, agents can change external systems.

1. Meaning and working



An agent interprets objectives, retrieves context, selects application programming interfaces (APIs), executes permitted steps, tests outcomes and replans after failures. Its autonomy depends on permissions, task boundaries and human approval.

2. Applications with examples

- **Business:** Microsoft Copilot Studio agents coordinate enterprise workflows; an illustrative grievance agent could route complaints to officials.
- **Science and software:** DeepMind's AlphaEvolve tests candidate algorithms; Claude Code assists with multi-step coding tasks.
- **Personal services:** OpenAI's Operator navigated websites and performed authorised browser actions.
- **India:** DST supported commercialisation of Fixit, an indigenous enterprise agent platform, in August 2026.

3. Advantages, risks and challenges

- **Advantages:** Continuous, adaptable workflows can reduce repetitive coordination, accelerate service delivery and support skilled workers.
- **Security:** Prompt injection may hijack goals; excessive permissions enable tool misuse, unauthorised transactions and data leakage.
- **Reliability:** Hallucinations, biased outputs and cascading errors can make consequential decisions unsafe.
- **Implementation:** Legacy-system integration, computing costs, interoperability, robust evaluation, privacy compliance and workforce reskilling complicate deployment. Independent testing and secure identity management are essential before granting operational authority.

4. Human oversight and accountability

Use least-privilege tools, action limits, audit trails and human approval for high-impact decisions. Deploying institutions retain responsibility and provide remedies.

Agentic AI delivers sustainable gains only when useful autonomy is matched by verifiable outcomes, enforceable boundaries and accountable human control.

What is agentic Artificial Intelligence (AI) ? Explain its working. Describe its applications with suitable examples. Discuss the advantages, risks and challenges associated with agentic AI systems.

DEMAND
Define agentic AI; explain goal-directed working; describe applications with concrete examples; discuss advantages, risks and implementation challenges

HIDDEN DEMAND
Autonomous actions need risk-sensitive human control, bounded permissions, institutional accountability and remedies

STRUCTURE
About three-tenths each for meaning and working, applications, and advantages with risks and challenges; about one-tenth for human oversight and accountability

BALANCE
Introduction 25 words, conclusion 20 words; 251 words in all

Why it works: The introduction distinguishes agentic AI from response-only generative AI while covering its capability to act. Part 1 explains the goal-to-feedback cycle and clarifies that autonomy is bounded. Part 2 provides four application categories supported by documented examples, including Microsoft's enterprise agents, AlphaEvolve, Claude Code, Operator and India's Fixit commercialisation initiative; the grievance example is explicitly illustrative. Part 3 balances conditional productivity gains against the distinctive risks of autonomous execution, including prompt injection, permission abuse, cascading errors and deployment constraints. Part 4 answers the hidden demand through least-privilege permissions, human approval, auditability and institutional responsibility. The conclusion gives a clear, balanced judgment: autonomous capability is valuable only with reliable outcomes and accountable control.

6 Other ways to write it

The model answer above remains the answer. These outlines show how different, equally valid answers can be constructed without losing any of the question's four demands. The choice depends on whether the student communicates best through a process, comparative assessment or concrete case.

Approach	How it is built	When it works
Goal-to-action process	Define agentic AI; draw the objective, planning, action and feedback cycle; give four applications; assess benefits and risks; close with human oversight.	**Best for this question.** It directly demonstrates the working mechanism while preserving applications and balanced discussion.
Application-risk matrix	Definition and working first; then a small table connecting enterprise workflows, software, scientific discovery and personal assistance to examples, benefits and corresponding risks.	When the student wants every application to demonstrate both opportunity and vulnerability.
Capability-based answer	Define agency through planning, tool use, feedback and bounded autonomy. Use each capability to explain relevant applications and the new problems it creates.	When the distinction between ordinary AI and agentic AI is the strongest organising idea.
Indian development perspective	Define and explain the technology; use Fixit, India's enterprise-agent startups and global examples; discuss productivity, India's technology-services opportunity, cybersecurity and governance.	When Indian policy relevance is important. Keep the global examples because the question does not restrict applications to India.
Trustworthy-autonomy approach	Explain the technology and three application categories; organise the assessment around reliability, security, privacy, accountability and operational controls.	When the student has a strong command of AI safety, but must still reserve adequate space for advantages and examples.
Human-agent collaboration	Start with the move from assistance to action; explain the workflow; contrast automation gains with professional judgment and employment transition; propose risk-sensitive human control.	When highlighting the hidden demand on responsibility. Avoid allowing governance to displace the three explicit demands.

What is agentic Artificial Intelligence (AI) ? Explain its working. Describe its applications with suitable examples. Discuss the advantages, risks and challenges associated with agentic AI systems.

7 Quick recall and Prelims crossover

Sixteen facts and distinctions for rapid revision. The technological concepts support the Mains explanation; the dates, organisations, policy instruments and named systems are useful for Prelims and for making examples precise.

1. **Agentic AI**: goal-directed AI with relative autonomy, planning, tool-enabled action and feedback. It need not operate without supervision.
2. **Core working cycle**: objective, observation, planning, action, feedback and possible replanning. There is no single mandatory architecture.
3. **Tool calling**: an agent requests operations through APIs, browsers or software functions. A proposed action is not automatically a successful action.
4. **IndiaAI Mission**: Union Cabinet approval on **7 March 2024**; an initiative for the wider AI ecosystem, not only autonomous agents.
5. **₹10,371.92 crore**: approved IndiaAI Mission outlay, across **seven** programme pillars, including Safe & Trusted AI.
6. **India AI Governance Guidelines**: released **5 November 2025**; **seven Sutras** and **six governance pillars**; a policy framework, not an agent-specific Act.
7. **DPDP Rules, 2025**: notification dated **13 November 2025**, Gazette publication **14 November**, with phased commencement.
8. **NITI Aayog roadmap**: **February 2026**, *India's Technology Services: Reimagination Ahead*; proposes an Agentic AI play and hybrid human-agent models.
9. **Operator**: OpenAI introduced the browser-using agent on **23 January 2025** as a research preview.
10. **ChatGPT agent**: introduced by OpenAI on **17 July 2025**, combining research with computer-use and task execution.
11. **Claude Code**: introduced by Anthropic on **24 February 2025** as a limited research preview for agentic coding.
12. **AlphaEvolve**: Google DeepMind's algorithm-discovery agent, announced **14 May 2025**; combines model-generated candidates with automated evaluation.
13. **OWASP Agentic Top 10**: released **9 December 2025**; addresses risks including goal hijacking, tool misuse and privilege abuse.
14. **NIST AI Agent Standards Initiative**: announced **17 February 2026**; focuses on secure agent operation, interoperability and standards.
15. **IndiaAI Startups Global**: **10 Indian startups** reported selected on **25 March 2026**, including voice-agent platform Voicing AI.
16. **Fixit**: TDB-DST announced support for One2X Tech's indigenous agentic AI platform on **20 August 2026**, for commercialisation.

THE ONE SENTENCE TO CARRY OUT OF THIS DOCUMENT

Agentic AI turns intelligence into action; its value depends on what the system can do, what it is permitted to do, and who remains accountable when it acts.

What is agentic Artificial Intelligence (AI) ? Explain its working. Describe its applications with suitable examples. Discuss the advantages, risks and challenges associated with agentic AI systems.

8 Write it yourself

Write the answer within the UPSC limit of 250 words without looking back, then check it against the seven requirements below. Count every whitespace-separated token, including section numerals, headings, point labels and text inside any flow or table.

	Check	Yes / No
1	Does the introduction **define agentic AI** through goal-directed autonomy and the capacity to act, rather than generative capability alone?	
2	Does Part 1 **explain the complete working cycle**, including goal, planning, tool use, action, feedback and the limits of autonomy?	
3	Does Part 2 cover **administrative or business, scientific or technical, and personal or service applications** with examples that are genuinely agentic?	
4	Does Part 3 explain **advantages and productivity gains** without treating possible benefits as established universal outcomes?	
5	Does the same part cover **security risks, reliability problems, accountability and implementation challenges**, rather than discussing privacy alone?	
6	Does the small final part establish **human oversight, bounded authority and institutional accountability**, with a conclusion that gives a clear judgment?	
7	Is the answer **250 words or a little over** under the stated house-counting rule, with accurate examples and none of the incorrect claims in section 4.2?	

Sources

- **Union Cabinet and Press Information Bureau:** IndiaAI Mission approval, **7 March 2024**, including the **₹10,371.92 crore** outlay and programme components. pib.gov.in. Ministry of Electronics and Information Technology: parliamentary update on mission progress, **27 July 2026**, including development of domain-specific AI systems. pib.gov.in.
- **Ministry of Electronics and Information Technology:** *India AI Governance Guidelines*, released **5 November 2025**, with seven guiding Sutras, six governance pillars and an action plan. Primary government report: static.pib.gov.in. Government's explanation of their legal and regulatory status, including reliance on existing laws and sectoral oversight: pib.gov.in.
- **Ministry of Electronics and Information Technology and Gazette of India:** Digital Personal Data Protection Rules, 2025, G.S.R. 846(E), dated **13 November 2025**, published in the Gazette on **14 November 2025**, including the staggered commencement provisions in Rule 1. Primary notification: meity.gov.in. PIB explanatory backgrounder of **17 November 2025**: pib.gov.in.
- **NITI Aayog, Frontier Tech Hub:** *India's Technology Services: Reimagination Ahead*, **February 2026**. The report identifies an **Agentic AI play**, productised services, legacy-system modernisation, hybrid human-agent models, and infrastructure, talent, R&D and policy enablers. Its projections and opportunities are forward-looking, not observed economic outcomes. niti.gov.in.
- **Ministry of Science and Technology, Technology Development Board:** announcement dated **20 August 2026** of commercialisation support for **One2X Tech Pvt. Ltd.** and its indigenous agentic AI platform **Fixit**. The release describes enterprise orchestration, integrations, security and scalability as project objectives. pib.gov.in.
- **Ministry of Electronics and Information Technology, parliamentary reply:** **25 March 2026** statement regarding ten selected startups under **IndiaAI Startups Global**, with Voicing AI listed as an agentic AI platform for enterprise-grade intelligent voice agents. pib.gov.in.
- **Ministry of Electronics and Information Technology:** post-event report dated **2 March 2026** confirming that the **India AI Impact Summit 2026** was held in New Delhi from **16 to 21 February 2026**, with a focus on responsible AI and developmental impact. pib.gov.in.

What is agentic Artificial Intelligence (AI) ? Explain its working. Describe its applications with suitable examples. Discuss the advantages, risks and challenges associated with agentic AI systems.

- **Microsoft:** *Unlocking autonomous agent capabilities with Microsoft Copilot Studio*, published **21 October 2024**. Describes autonomous-agent capabilities, enterprise workflow orchestration and public-preview plans for Microsoft Ignite 2024; includes Pets at Home among the business examples. microsoft.com.
- **OpenAI:** *Introducing Operator*, **23 January 2025**, documenting a browser-using AI agent able to click, type and scroll as part of task execution. openai.com. *Computer-Using Agent*, also **23 January 2025**, explaining the underlying interaction model and its evaluations: openai.com. *Introducing ChatGPT agent: bridging research and action*, **17 July 2025**: openai.com.
- **Anthropic:** *Claude 3.7 Sonnet and Claude Code*, **24 February 2025**. Introduces Claude Code as a command-line tool for agentic coding, initially available as a limited research preview. anthropic.com.
- **Google DeepMind:** *AlphaEvolve: A Gemini-powered coding agent for designing advanced algorithms*, **14 May 2025**. Explains the combination of language models, evolutionary search and automated evaluation for algorithm discovery and optimisation. deepmind.google.
- **OWASP GenAI Security Project:** *Top 10 for Agentic Applications 2026*, released **9 December 2025**, identifying security risks in autonomous agents and practical mitigations. Official resource: genai.owasp.org. Accompanying explanations of agent goal hijacking, tool misuse, privilege abuse, memory poisoning and other threats: genai.owasp.org.
- **United States National Institute of Standards and Technology:** *Announcing the AI Agent Standards Initiative for Interoperable and Secure Innovation*, **17 February 2026**. Describes a standards-development initiative addressing secure autonomous operation and interoperable agent systems. nist.gov.
- **Union Public Service Commission:** Civil Services (Main) Examination, official previous question papers, including General Studies Paper III, 2026, and the earlier GS papers quoted in section 2.2. upsc.gov.in. Examination-specific 2026 records and question-paper listings: cms.upsc.gov.in.

What is agentic Artificial Intelligence (AI) ? Explain its working. Describe its applications with suitable examples. Discuss the advantages, risks and challenges associated with agentic AI systems.

YOUR NEXT STEP

Now write GS Paper III 2026 again.

You now know what this question asked, what it expected without saying so, and how its answer is built. Write the whole paper under the clock, have it evaluated, and read the feedback against this document.



1 Write the full paper in three hours

2 Self-check this answer with section 8

3 Have the paper evaluated

4 Read the feedback against this document

FREE

Mains Replay

170 UPSC Mains papers from 2017 to 2026 (GS I to IV, Essay and six optionals), with a built-in three-hour exam clock. Write GS Paper III 2026 exactly as you would in the hall.

upscgurus.in/mains-pyq

₹1,499 PER PAPER, INCL. GST

Paper Evaluation

A written evaluation of every answer and a personal feedback session, through Mains Mastery.

upscgurus.in/mains-mastery

FREE

UG Handout 01

Where did my Mains 2026 marks go? All 79 GS questions and parts decoded: demand, hidden demand, structure.

upscgurus.in/open-library

FREE

GS Mains PYQ Deep Decode

Previous year questions taught on YouTube: the reasoning, the structure and what a strong answer holds.

UPSC Gurus on YouTube

PROGRAMME

Mains Mastery

The UPSC Gurus GS Mains programme: every question read for its demand, hidden demand and structure, and answers written and evaluated to that standard.

upscgurus.in/mains-mastery

UPSC Gurus is an inclusive platform. This PDF is tagged, with headings, reading order and selectable text, for use with screen readers.

Guranchal Ma'am

Founder, UPSC Gurus