Report 2: Agentic AI: Autonomous Decision-Making Systems

Agentic AI refers to AI systems capable of independent action, decision-making, and goal pursuit in dynamic environments. These systems integrate LLMs' language abilities with planning, perception, and tool use.

Core Concepts

- Agentic AI exhibits autonomy, proactively setting and pursuing objectives.

- They use multi-step reasoning, planning, and reflection to adapt strategies.

- Memory modules allow agents to retain long-term context for continuity and learning.

- Interaction with external tools and APIs enables expanding agent capabilities beyond language understanding.

Architectures and Design

- Agentic AI often operates iterative loops of action generation, evaluation, reflection, and adjustment.

- Multi-agent systems allow collaboration, competition, or coordination for complex problem-solving.

- Safety considerations include sandboxed environments, robust failure modes, and ethical alignment.

Use Cases

- Digital secretaries managing workflows and communications.

- Research agents automating literature review and experiment planning.

- Customer service bots with autonomy to resolve complex requests.

- Autonomous robotic systems integrated with LLMs.

Challenges

- Balancing autonomy with controllability.

- Mitigating errors, biases, and unintended behaviors.

- Designing interpretable decision-making processes.

- Ensuring alignment with human values and ethics.