Agentic AI: Introduction

What is an autonomous agent?

- An autonomous agent is an AI system that can perform tasks and make decisions on its own, without human intervention.
- Traditional definition:
  - An agent is an entity that perceives its environment through sensors and acts upon that environment through actuators.
- Many different types of agents, including:
  - Software agents: chatbots, virtual assistants, recommendation systems
  - Physical agents: robots, self-driving cars, drones
- In the last year, the term "agent" has become popular in the context of large language models (LLMs) and AI systems that can perform complex tasks by themselves.

The problem: perform research for a hardware project

- I want to build a hardware project, for instance to count the number of deer in my backyard
- I have no experience with hardware, and I don't know where to start
- I want to use an AI agent to help me with this task, by doing research, finding resources, and guiding me through the process

What makes an agent problem?

- **Inputs**:
  - A task or goal that the agent needs to accomplish (eg. "count the number of deer in my backyard")
  - Possibly, some constraints or preferences (eg. "I want to use open source hardware", "I want to spend less than $1000")
- **Outputs**:
  - Research artifacts: design diagrams, code snippets, lists of materials, step-by-step instructions, etc.
  - Collaboration with the user: asking questions, providing feedback, suggesting alternatives, etc.
  - Collaboration with other agents
  - A final product that accomplishes the task (eg. a working deer counting system)

How do I know that it works?

- **Performance measure**:
  - User satisfaction: does the user feel that the agent helped them achieve their goal?
  - Task completion: did the agent successfully accomplish the task (eg. did it help build a working deer counting system)?
  - Quality of the outputs: are the research artifacts useful, accurate, and relevant to the task?
  - Efficiency: how much time and resources did it take for the agent to accomplish the task?

Let us hack a solution!

- I can make a **plan** for what I would need:
    i. **Research the problem**: find out what is needed to build a deer counting system (eg. sensors, microcontrollers, software, etc.)
    ii. **Find resources**: look for tutorials, guides, and examples of similar projects
    iii. **Design the system**: create a design for how the components will work together
    iv. **Build the system**: acquire the materials and assemble the hardware
    v. **Test and iterate**: test the system and make improvements as needed

Let us hack a solution! (continued)

- Some of the components of the plan are based on AI skills we have already seen:
- **Object detection**: to count the number of deer, we need to be able to detect them in images or video
- **Summarization**: to research the problem, we need to be able to summarize information from various sources
- and so on...
- The trick is to combine them

Agentic AI

- The term "agentic AI" is often used to refer to AI systems that have the ability to perform tasks autonomously, without human intervention.
- This can include a wide range of capabilities, such as:
  - Planning and decision making: the ability to create a plan to achieve a goal and make decisions based on that plan
  - Learning and adaptation: the ability to learn from experience and adapt to new situations
  - Collaboration: the ability to work with humans and other agents to accomplish tasks
  - Creativity: the ability to generate novel ideas and solutions to problems

Example: Codex

- Codex is a language model developed by OpenAI that can generate code from natural language descriptions
- It can be used as an agent to help with programming tasks, such as:
  - Writing code snippets
  - Debugging code
  - Generating documentation
  - Collaborating with the user to solve programming problems
- It can also be used in combination with other tools and agents to accomplish more complex tasks, such as building a hardware project from scratch.

Other examples of agentic AI

- Claude Code
- Cursor
- Github Copilot
- Many of these examples are focusing on programming tasks.

OpenClaw (formerly Clawdbot, Moltbot, etc.)

- OpenClaw is an open source project that aims to create a general-purpose agentic AI system that can perform a wide range of tasks, including research and hardware projects.
- It is designed to be modular and extensible, allowing users to customize and extend its capabilities

Pitfalls and dangers: Security

- In order to build the deer counting project, I will need to grant extensive rights to the agent:
  - Access to my Amazon account to buy things and ship it to my home
  - Access to my bank account to pay for them
  - When installed, the agent will have access to my home network and all the devices connected to it
  - And it has access to a camera in my backyard, which could be used to spy on me and my family, and my neighbors

Pitfalls and dangers: Job displacement

- One of the main concerns with autonomous agents is that they could displace human workers in various industries.
- For instance, if we have an agent that can perform research and build hardware projects, it could potentially replace human researchers and engineers.
- This could lead to job losses and economic disruption, especially for those who rely on these jobs for their livelihood.
- It is important to consider the ethical implications of autonomous agents and to find ways to mitigate these risks, such as retraining workers for new roles that complement the capabilities of agents rather than compete with them.

Try it out: