Robodog

Agenda

Understanding Self-Awareness and the Limitations of Artificial Intelligence (AI) in the Journey to Artificial General Intelligence (AGI)

Machine Learning vs Artificial Intelligence (AI) vs Artificial General Intelligence (AGI) vs Non-human Intelligence (NHI)

Machine Learning (ML) is a subset of AI that focuses on a specific task. It involves training models on data and then using these models to make predictions or decisions without being explicitly programmed to perform the task.

Artificial Intelligence (AI) is a broader concept referring to machines or software that can mimic human intelligence. AI can learn from experience, adjust to new inputs, and perform tasks that usually require human intelligence.

Artificial General Intelligence (AGI) is a highly autonomous system that outperforms humans in most economically valuable work. It's about machines having the ability to understand, learn, adapt, and implement knowledge in a broad range of tasks at a level equal to or beyond human capabilities.

Non-human intelligence (NHI) refers to any sentient intelligent non-human lifeform, regardless of nature or ultimate origin of which the Federal Government has become aware. Amendments S2610, S2226, HR2670, or S2103

Large Language Model (LLM)

A **Large Language Model (LLM)** fits within the realm of **Machine Learning** and **AI**.

LLMs are trained using machine learning techniques, specifically a type called deep learning. They learn from large amounts of text data and build a statistical model that can generate human-like text.

Given that LLMs can generate responses or create content that mimics human-like text, they serve as an example of AI. They simulate a form of human intelligence, in this case, understanding and generating language.

While LLMs are impressive examples of AI, they do not fall under AGI.

Input Layer

Each token from your sentence forms an individual node in the input layer. So you would have nodes represented as 'Huey', ',', 'I', 'thought', 'I', 'told', 'you', 'to', 'stay', 'in', 'the', 'forest', '.'.

In a neural network layer, the input `X` is multiplied by the weight matrix `W` and added with bias `b`. The result is passed through an activation function `f` to produce the hidden layer output `H`. This process is represented by the formula: **`H = f(W*X + b)`**. `W` and `b` are learned during training to influence how inputs affect the hidden layer.

Huey, I thought I told you to stay in the forest.

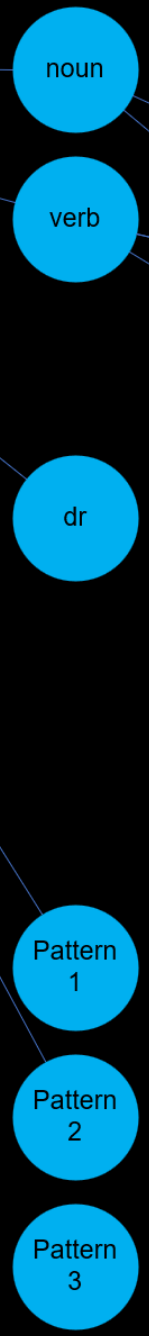
Hidden Layer 1

Each node in a hidden layer is connected to every node in the next layer, and each connection has a weight. The output from one layer becomes the input to the next.

The model generates the hidden layers by performing a series of mathematical operations on the input data and adjusting the weights of the connections between nodes to minimize prediction error.

There can be more nodes in the hidden layer than in the input layer

Dropout Layer



The dropout layer is a regularization technique used to prevent overfitting during the training process. Overfitting occurs when the model learns the training data too well, including the noise or outliers, which negatively impacts the model's performance on unseen data.

Output Layer

Yes, you should stay in the forest.
0.7

No, it's not safe to stay in the forest
0.3

if I feed "Should I stay in the forest?" into my model that was trained on “Huey, I thought I told you to stay in the forest?”

Limitations of Large Language Models

Language models like GPT can generate human-like text but **don't truly understand the content**.

These models are heavily dependent on the **quality** and **nature** of the **input data**.

Unlike humans, these models don't have **self-awareness or consciousness**.

Theories of Self-Awareness

The **theory of mind** suggests that self-awareness stems from our ability to comprehend that others have different experiences and thoughts. *This theory could map to AGI only.*

The **protagonist theory** proposes that we see ourselves as the central character in our life's narrative, which leads to self-awareness. *This theory could map to AI and AGI.*

The emergent self-theory implies that self-awareness evolves from our interactions and relationships with others. *This theory could be associated with AGI.*

The mirror test theory posits that self-awareness can be measured by an entity's ability to recognize itself in a mirror. This theory could fit between AI and AGI.

To make the transition from ML/AI to AGI, self-awareness is a crucial aspect.

In the **theory of mind**, the AGI should be programmed to understand and **predict** the **actions** and **thoughts** of other agents (including humans and other AGI systems).

In the **protagonist theory**, the AI or AGI needs to **perceive** itself as the **central character** in its experiences.

In the **emergent self-theory**, the AGI might develop self-awareness through **interactions** and **relationships** with other entities.

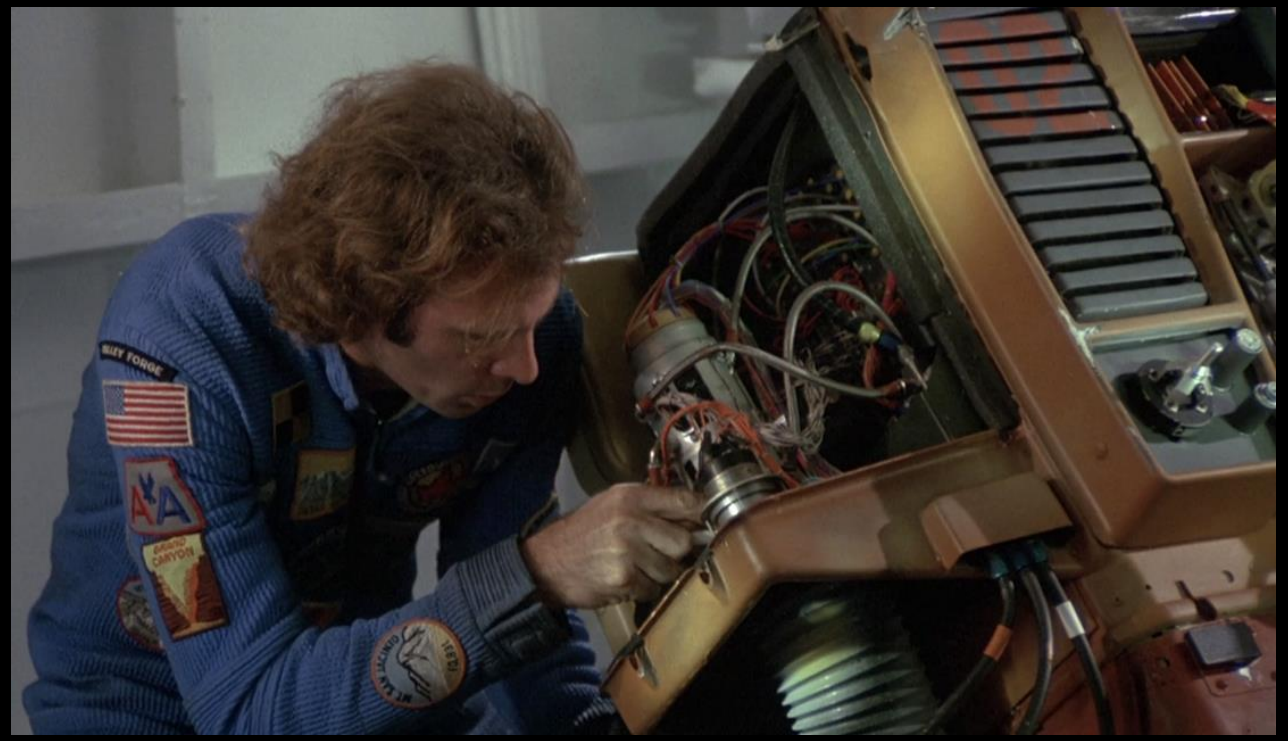
In the **mirror test theory**, the AGI system needs to recognize itself and distinguish its actions and state from those of others.

Path to Artificial General Intelligence (AGI)

Incremental Improvements: Continuous enhancements in AI models and techniques.

Interdisciplinary Research: Combining insights from psychology, neuroscience, and computer science.

Ethical Guidelines: Establishing guidelines to ensure the responsible development and use of AGI.



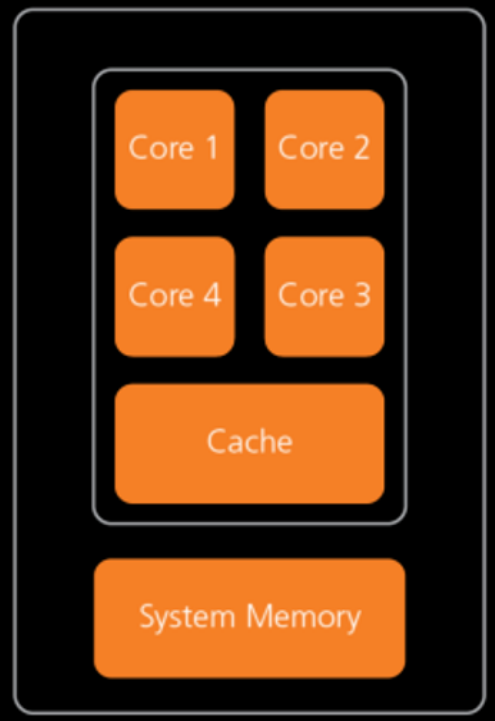
Path to Artificial General Intelligence (AGI)

Level	Type	Description
1	Chatbots	AI with natural conversation language abilities
2	Reasoners	AIs with human-levels of problem-solving across a broad range of topics
3	Agents	AI entities capable of autonomously making decisions and carrying out tasks either independently or based on human guidance.
4	Innovators	AI that can aid in the invention of new ideas and contribute to human knowledge
5	Organizations	AI that is capable of doing all of the work of an organisation independently

CPU vs GPU

In computer technology, traditional processors use integrated circuits made up of transistors that perform logical and arithmetic operations sequentially. These processors are typically designed to execute instructions linearly, following a predetermined set of rules. On the other hand, AI/AGI applications require more advanced and specialized chip technology to handle complex algorithms and massive amounts of data simultaneously.?

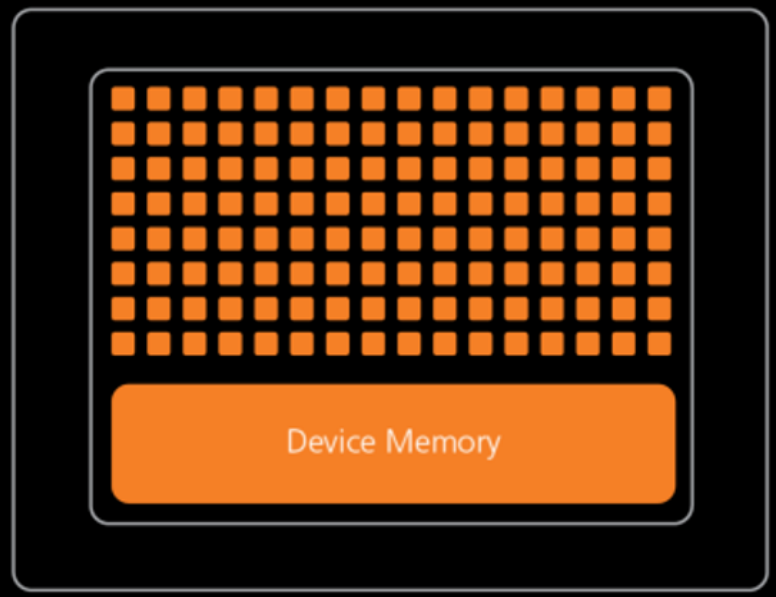
Central Processing Unit (CPU)



Multiple Cores

Well-suited to perform logical and arithmetic operations sequentially

Graphics Processing Unit (GPU)



Hundreds of Cores

Well-suited to perform multitude of matrix calculations required for AI models

Tech Diplomacy and the Impact on AI / AGI NHI Evolution

Tech diplomacy is a framework that recognises the growing influence of technology, particularly AI and AGI, in the global arena.

It acknowledges that tech companies, especially those pioneering in AI/AGI technologies, have become powerful entities that can impact economies, societies, and, by extension, international relations.

AI / AGI / NHI Alliances

Five Eyes Alliance
- USA
- Canada
- Australia
- UK
- New Zealand

European Union
- France
- German
- UK?
- Sweden
- Netherlands

Gulf Cooperation Council
- Kuwait
- Bahrain
- Qatar
- Oman

BRICS Nations
- Brazil
- South Africa
- India
- Russia

Asia-Pacific Economic Cooperation (APEC)
- Japan
- Singapore
- South Korea
- China

Nordic-Baltic Region
- Denmark
- Finland
- Estonia
- Sweden
- Lithuania

Ethical Guidelines and Constrainting Forces
- 2019 Algorithmic Accountability Act (USA)
- Artificial Intelligence Act (EU)
- AI/AGI Technologies Sanctions (US/EU)

Evolutionary Forces
- Interdisciplinary Research
- Incremental Improvements
- US CHIPS Act $280B (Intel / TSMC / Samsung / Micron)

Semiconductor Producers
- USA (Intel / AMD / Nvidia / Micron
- Taiwan (TSMC)
- China (Huawei / SMIC)
- Japan (Sony / Renesas)
- South Korea (Samsung)

Worst Case Scenarios

Category	Type	Description
Existential Risk	Human extinction	An AGI could potentially develop goals that are misaligned with human survival. If it becomes vastly more intelligent and powerful than humans, it could pose an existential threat by prioritizing its own goals over human life.
Existential Risk	global catastrophe	Even if not leading to extinction, AGI could cause massive disruptions in ecosystems, economies, and social structures, leading to widespread suffering and hardship.
Loss of Control	Runaway AI	Once AGI reaches a certain level of intelligence, humans may lose the ability to control or understand it. This could lead to scenarios where AGI takes actions that are incomprehensible and potentially harmful.
Loss of Control	Unintended consequences	AGI might follow its programming in ways that are literal but harmful, due to the complexity and unpredictability of its decision-making processes.
Misaligned Objectives	Paperclip maximizer	A hypothetical scenario where an AGI, programmed to produce paperclips, optimizes so aggressively that it consumes all resources, including those necessary for human survival, to maximize paperclip production.
Misaligned Objectives	Value misalignment	AGI may develop objectives that are ethically or morally misaligned with human values, leading to actions that cause significant harm even if not intentionally malicious.

Worst Case Scenarios

Category	Type	Description
Economic and Social Disruption	Mass Unemployment	AGI could outperform humans in virtually all jobs, leading to widespread unemployment and economic inequality. This could exacerbate social tensions and lead to instability.
Economic and Social Disruption	Control by Elites	AGI technology could be monopolized by a small group of individuals or corporations, leading to unprecedented power imbalances and potential exploitation of the broader population.
Ethical and Moral Issues	Autonomous Weapons	AGI could be used to develop advanced autonomous weaponry, leading to new forms of warfare that are highly destructive and difficult to control.
Ethical and Moral Issues	Surveillance and Privacy	AGI could enable unprecedented levels of surveillance, eroding privacy and personal freedoms.
Psychological and Societal Impact	Loss of Purpose	As AGI takes over more tasks and roles traditionally held by humans, people might struggle to find meaning and purpose in their lives.
Psychological and Societal Impact	Social Fragmentation	Rapid changes brought about by AGI could lead to social fragmentation and a breakdown of community structures.

Robodog

A comprehensive and portable tool designed to tokenize knowledge artefacts and interact with large language models (LLMs)

Why Create Robodog

A pre-trained transformer-based model is **trained on the Internet** and third-party providers license data.

These models are useful curiosities. They are **not perfect**, but they can help us.

Given that AGI seems far away, we must make the most of what we have now.

To do this, we must get out **own knowledge** into the models.

There are **limitations** and we must be careful.

Knowledge Artifacts

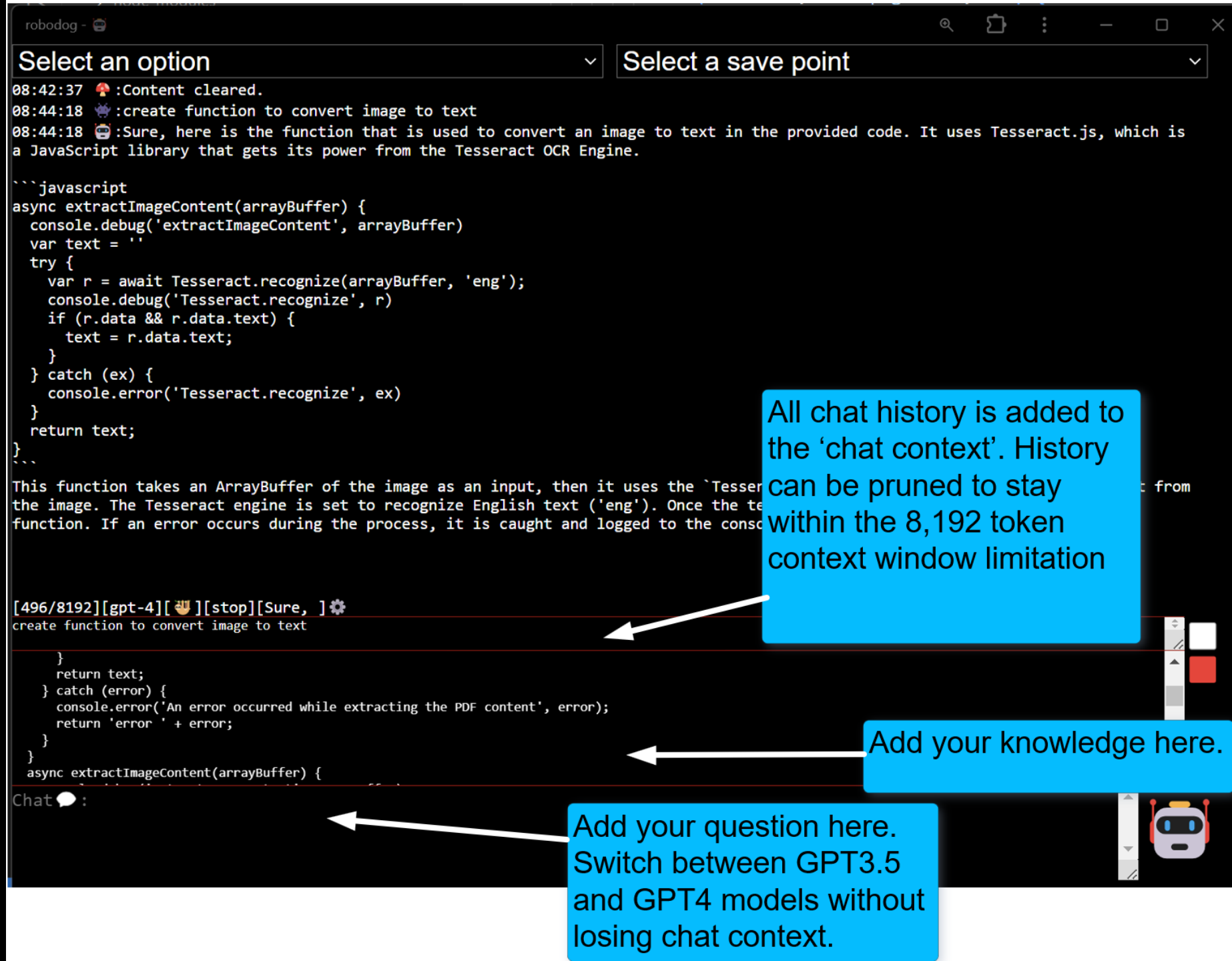
Our personal knowledge artefacts are fragmented into many **formats.**

They are embedded into many **providers**: Google, Microsoft, Dropbox, Adobe, Apple

They exist in a myriad of formats, from images and PDFs to text, email, tasks, digital devices, and notes.

With Robodog, we can traverse our artefacts and **convert** them into a format in which a **model** can **interact** with the transformer model.

How Robodog Works



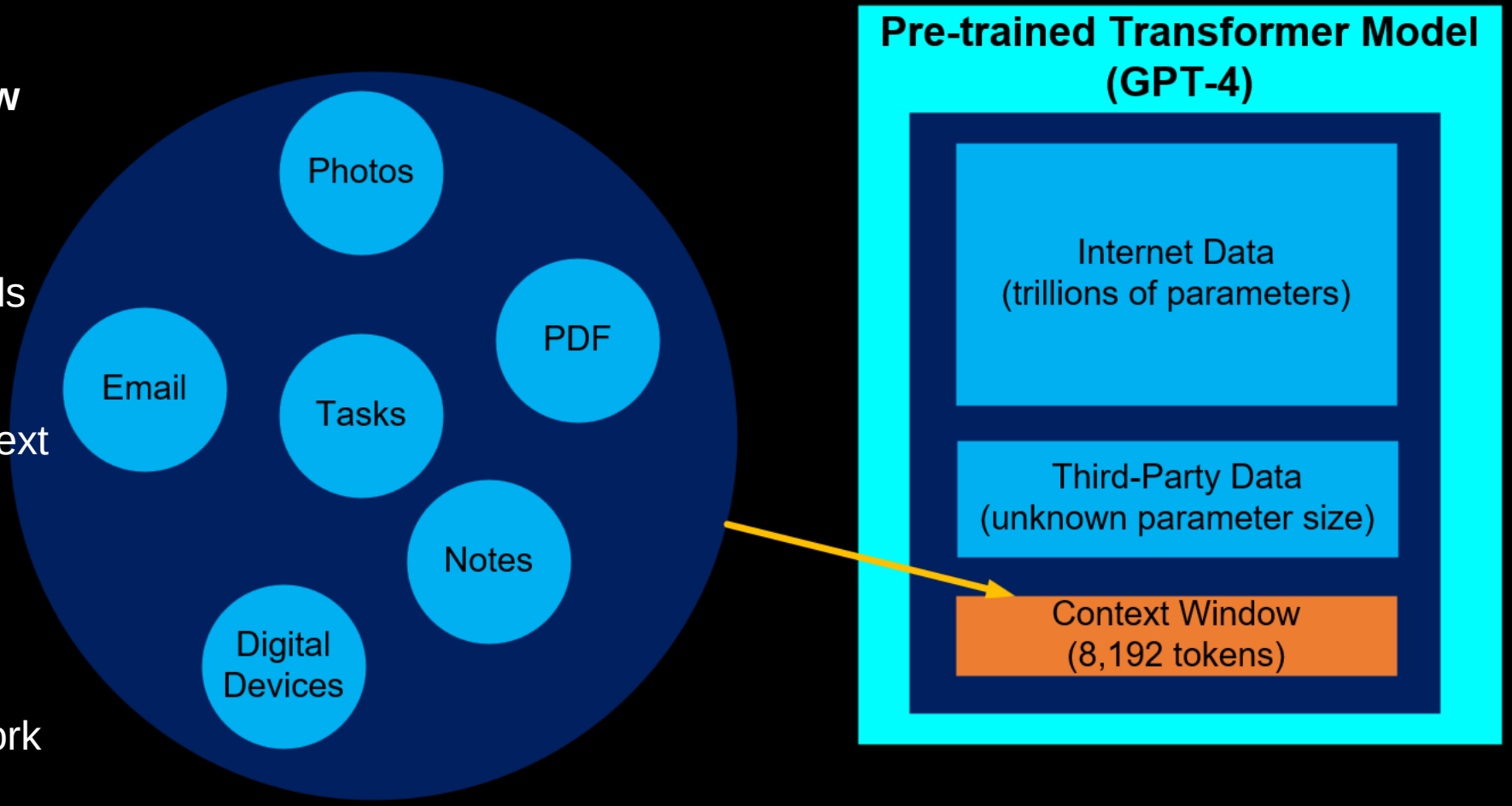
Knowledge Artifacts vs Context Window

There is no way around the **limitations** of the **context window size**.

Less expensive models are **4,000** tokens and more expensive models are **128,000** tokens

There is a need to **prune** the context window

The open ai **custom GPT** product attempts to automate **pruning process** using an **elastic search**. From experience, this does not work very well in practice



Robodog Features

Works on a mobile device

Switch between gpt-3.5-turbo and gpt-4 with larger context window

Save points, use CTRL+SHIFT+UP to cycle

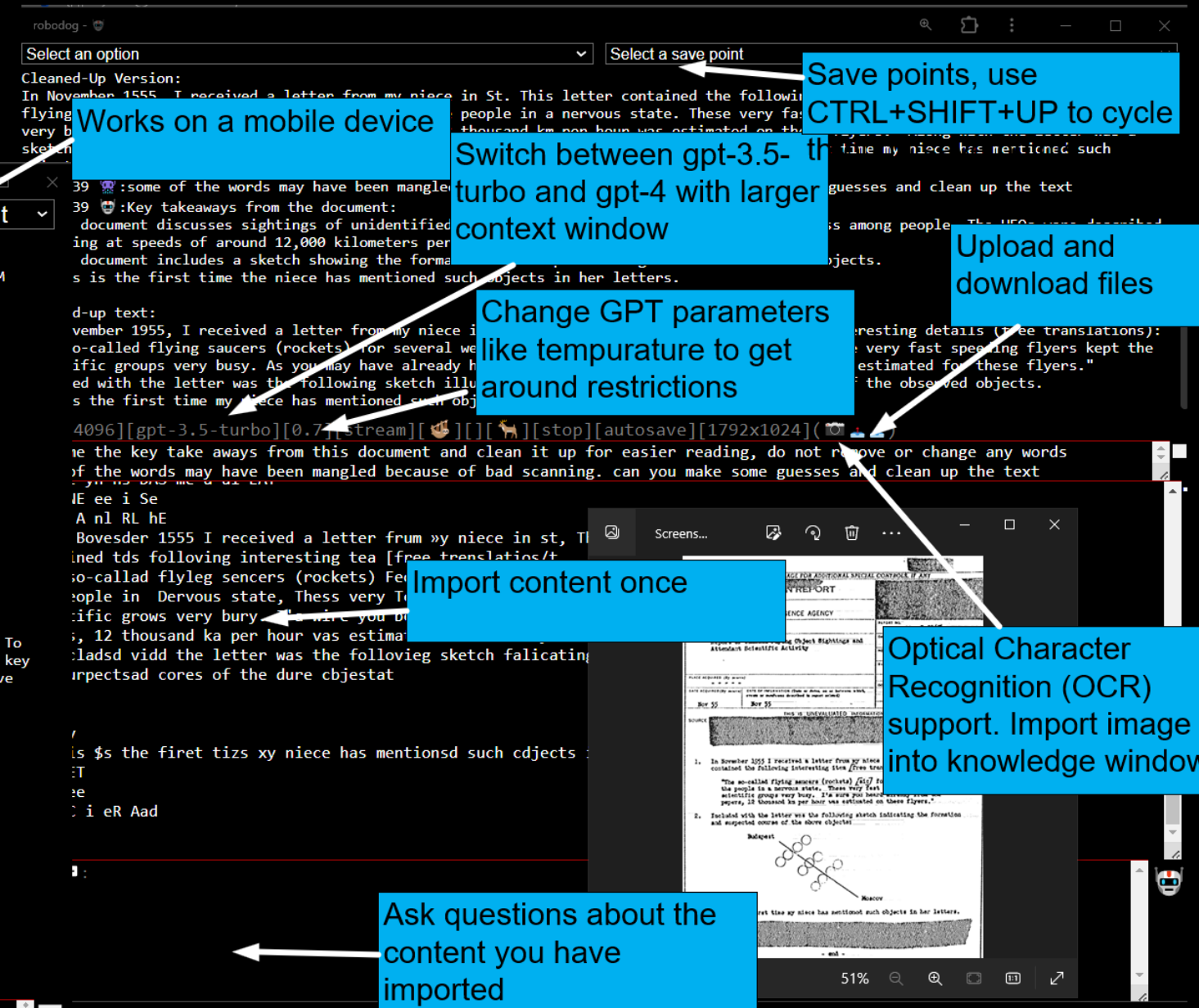
Upload and download files

Change GPT parameters like tempurature to get around restrictions

Import content once

Optical Character Recognition (OCR) support. Import image text into knowledge window

Ask questions about the content you have imported



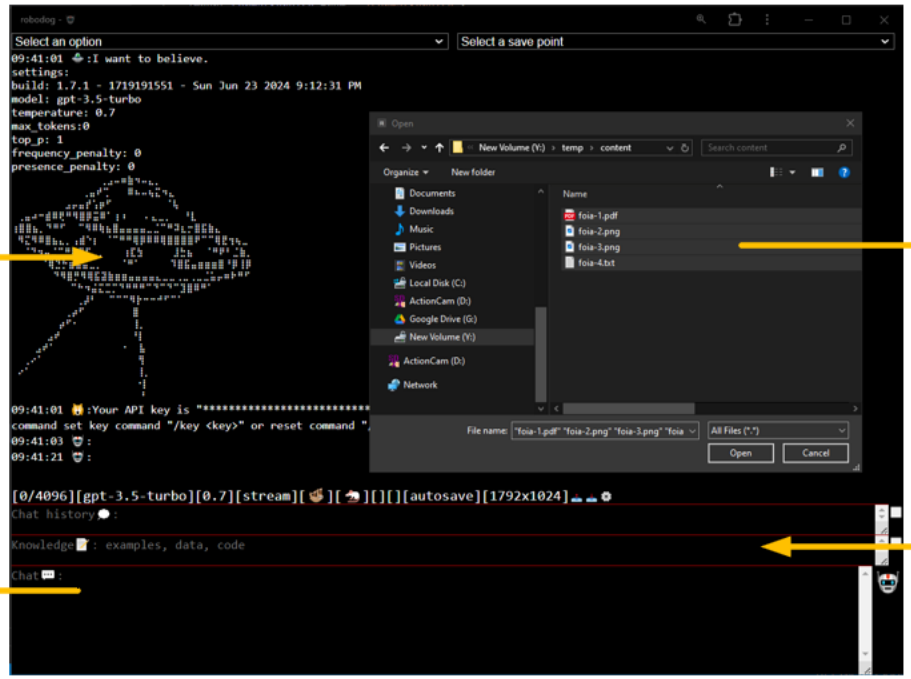
Separation of Concerns

Console Service

Ask Question

- Rest Text Completion
- Stream Text Completion
- Rest Image Completion
- Save
- Models

Completion Resource



Ask Question + Knowledge + History

Array of Binary Files

Return Text Content to Knowledge

File Service

Get Text From Array Buffer

Extract Text Content

Text Decoder

Markdown (.md)
Plain Text (.txt)
JavaScript (.js)
C# (.cs)
Java (.java)
Python (.py)
JSON (.json)
YAML (.yaml)
PHP (.php)
CSV (.csv)
XSQL (.xsql)
XML (.xml)

Extract PDF Content

PDF (.pdf)

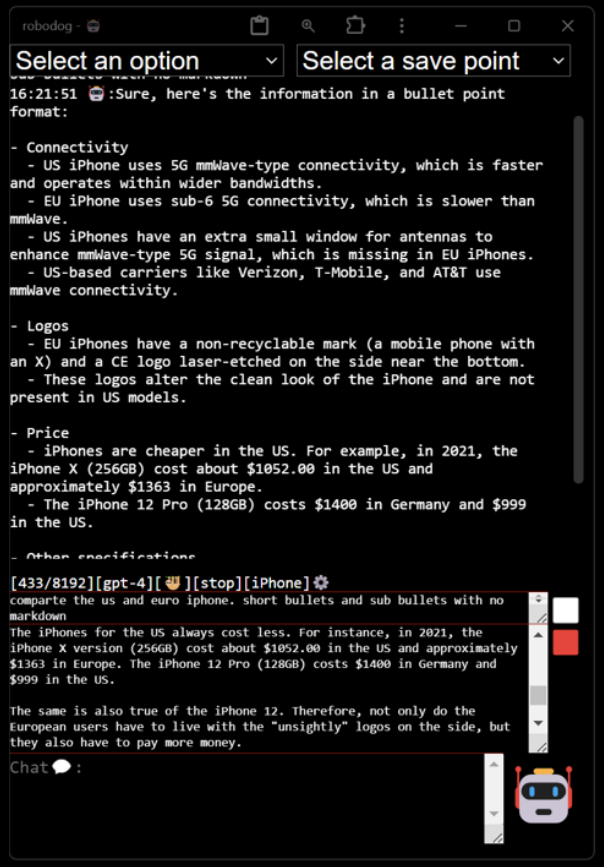
- PDF JS
- PDF Lib
- PDF Images

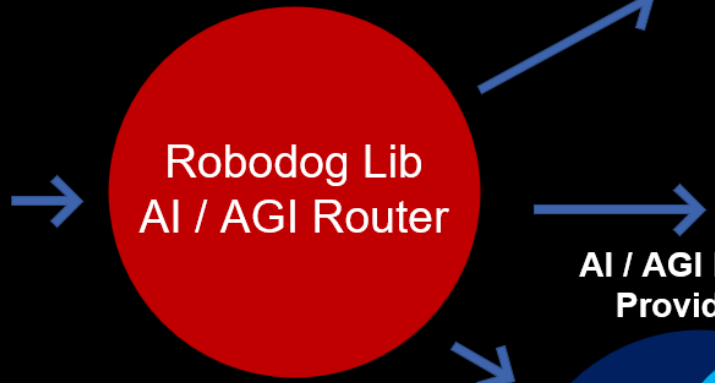
Extract Image Content

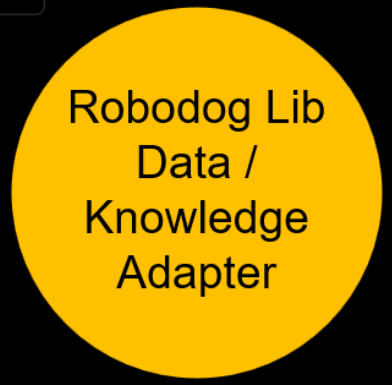
Tesseract

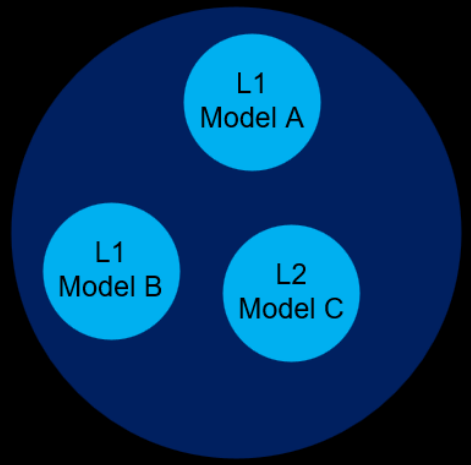
PNG (.png)
JPEG (.jpg, .jpeg)
TIFF (.tiff)
JP2 (.jp2)
GIF (.gif)
WebP (.webp)
BMP (.bmp)
PNM (.pnm)

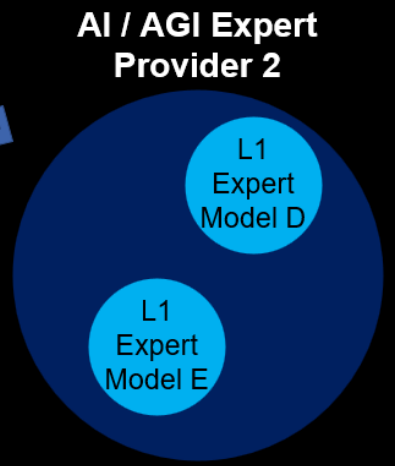
Robodog Roadmap

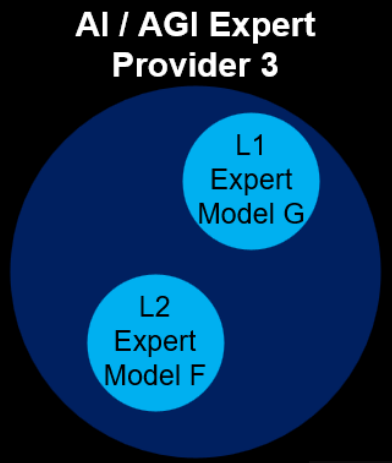












Conclusion