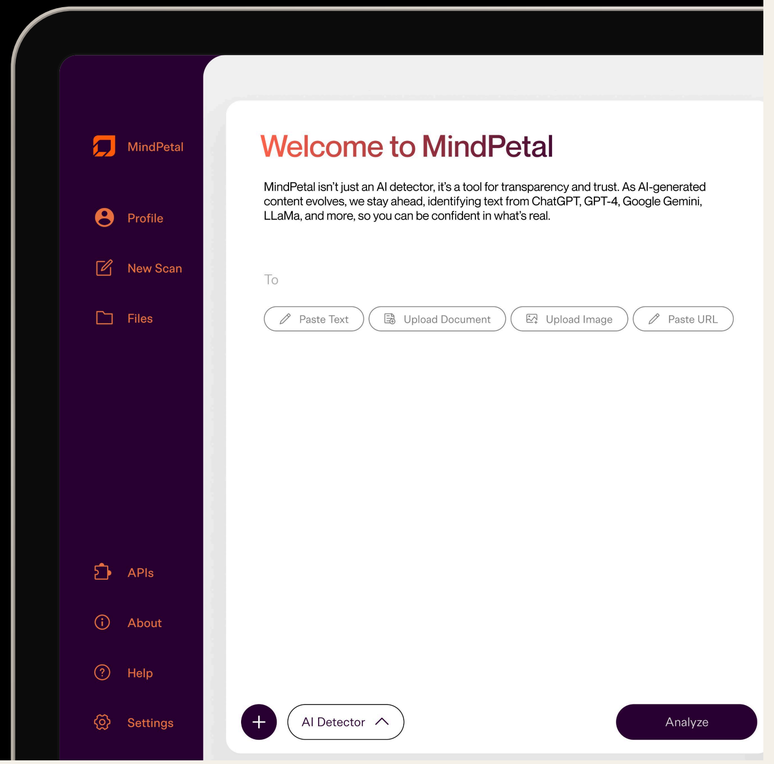
AI Recognition Tool Design

IC25

IC25082 - Sanjay Kumar Balaji, Samita Prakash Belliganood



Understanding the Challenge

The Problem: AI-generated content is harder to distinguish from human-created media.

The Goal:

- Design an AI content checker tool to verify authenticity and provide key metadata.
- Users can upload/input text, images, or other media to check if it's AI-generated.

Focus: Seamless user experience with intuitive and supportive features.

Our Design process:

- **Map:** Researched AI content checkers to understand how they work, their limitations, and key features to improve user experience. We also created personas to identify user needs and ensure our design is tailored to real-world use cases.
- **Sketch:** Explored a wide range of ideas through sketching and ideation exercises like Crazy 8s. We generated multiple low-fidelity wireframes to experiment with different layouts and interactions.
- **Decide:** Faced challenges in selecting a final design. Conducted A/B testing, sought external feedback, and debated different options before finalizing the best approach.
- **Prototype:** Developed a prototype based on our chosen design, refining user experience and ensuring smooth functionality.
- **Test (Next Step)**: Plan to test the prototype, gather feedback, and iterate to improve usability and effectiveness.

Map:
Researched AI content checkers to understand how they work, their limitations, and key features to improve user experience.

How does AI content detection work?

Many AI content detectors rely on the same techniques AI models like ChatGPT use to create language, including machine learning (ML) and natural language processing (NLP).

Machine learning (ML)

Machine learning is about recognizing patterns – the more text is analyzed, the more easily the tools can pick up the subtle differences between AI-generated and human-generated content. Machine learning drives predictive analysis, which is critical for measuring perplexity – a key indicator of AI-generated text.

Natural language processing (NLP)

Natural language processing is about the nuances of language, and helps AI detectors gauge the context and syntax of the text it is analyzing. AI can create grammatically correct sentences but tends to struggle with being creative, subtle, and adding depth of meaning (which humans naturally use in their writing).

Classifiers and Embeddings

Within these broad categories of ML and NLP, classifiers and embeddings play important roles in the detection process. Classifiers place text into groups depending on the patterns they have learned: much like teaching someone to sort fruits based on characteristics they've learned, but applied to language. Embeddings represent words or phrases as vectors, which create a 'map' of language – this allows AI detectors to analyze semantic coherence.

Perplexity

Perplexity is like a surprise meter when it comes to AI content detectors. The higher the perplexity, the more 'surprised' the detector is by the text it is seeing – as unexpected or unusual words or sentence structures tend to raise the perplexity score. If the text is ranked to have higher perplexity, it is more likely to be created by human. If the text is too predictable, it's likely to be AI authorship.

Burstiness

Burstiness is a measure of how much the perplexity varies over the entire document – and is more about how the text flows. While human writing has a rhythm of sho long phrases, mixing up both simple and complex sentences, AI can often be a lot more predictable. Which means its sentences tend to be fairly uniform.
This means AI generators often veer towards lower burstiness and create more monotonous text – repeating certain words or phrases too frequently because they' seen them appear often in their training data.

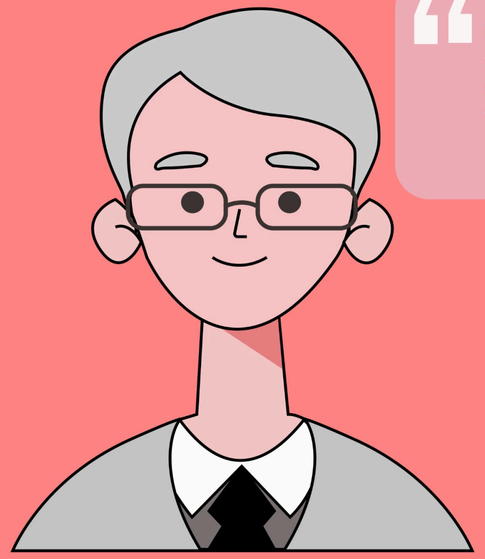
The interaction between perplexity and burstiness

While perplexity is about the individual surprises of specific words and phrases, burstiness is more about the overall rhythm of a piece.
A text with high burstiness can lead to higher perplexity as this is like a curveball to AI, with sudden shifts in sentence length making it harder for the AI to predict wha next. But low burstiness often means lower perplexity – uniform sentences mean that AI has an easier time spotting the pattern and predicting the next words.
AI content detectors look for a balance of perplexity and burstiness that mimic the way humans naturally write and speak. Too much of either perplexity or burstines red flag.

AI detectors vs. plagiarism checkers

While both AI detectors and plagiarism detection tools exist to verify the authenticity of content, they operate differently. AI detectors look at the text's structure a choice and overall style to see whether it was created by artificial intelligence or a human – and involves advanced algorithms and linguistic analysis.
Meanwhile, plagiarism checkers are more straightforward and are essentially looking to match the text – and compare the writing against a broad data set of existin When they spot similarities or matches, they will flag this as potential plagiarism.

Persona: Understanding Our Users



> I don't want to just catch AI use - I want to help students understand and improve their writing.

Professor James Carter

50, New York

Occupation: University Professor (English & Literature)

Education: State University of New York (SUNY)

Bio

Experienced English & Literature professor at SUNY with a focus on academic integrity and AI detection in student work. Skilled in AI content verification, digital learning tools, and student engagement. Regularly uses Turnitin, Google Docs, and Canvas LMS to assess originality but seeks a faster, more transparent AI detection tool with LMS integration and clear, student-friendly feedback. Passionate about guiding students in ethical writing while leveraging technology for fair assessments.

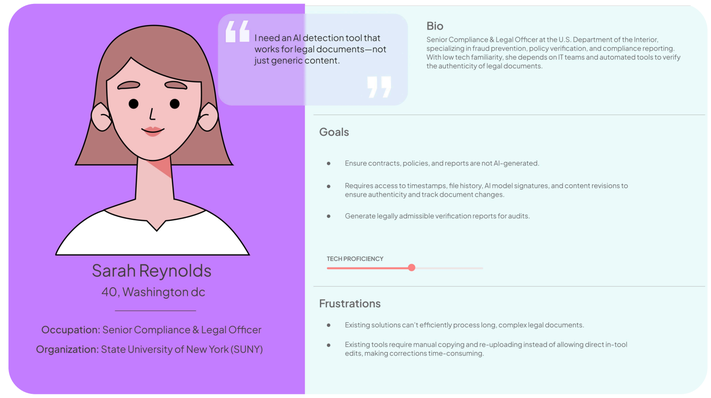
Goals

- Quick AI Content Detection: Needs a fast, reliable way to analyze student essays.
- Transparency in AI Score: Wants a clear breakdown of flagged content (not just an overall % score).
- Integration with LMS: Should be able to upload files from Canvas, Google Drive, or Dropbox easily.

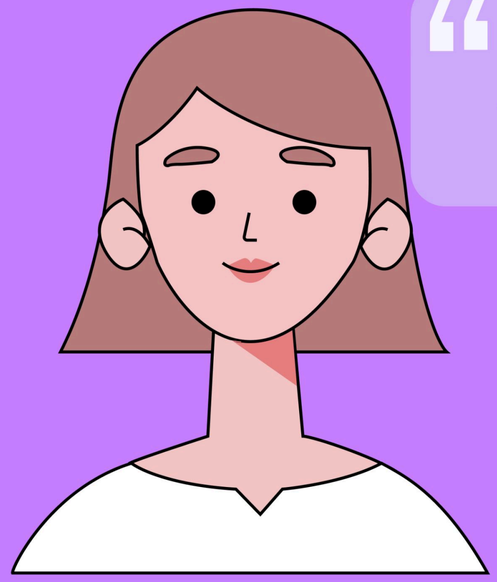
TECH PROFICIENCY

Frustrations

- AI-generated content is becoming harder to detect manually.
  Lack of Transparency: Some tools provide a probability score without explaining where a document is AI-generated.



Persona: Understanding Our Users



> I need an AI detection tool that works for legal documents—not just generic content.

Sarah Reynolds

40, Washington dc

Occupation: Senior Compliance & Legal Officer

Organization: State University of New York (SUNY)

Bio

Senior Compliance & Legal Officer at the U.S. Department of the Interior, specializing in fraud prevention, policy verification, and compliance reporting. With low tech familiarity, she depends on IT teams and automated tools to verify the authenticity of legal documents.

Goals

- Ensure contracts, policies, and reports are not AI-generated.
- Requires access to timestamps, file history, AI model signatures, and content revisions to ensure authenticity and track document changes.
- Generate legally admissible verification reports for audits.

TECH PROFICIENCY

Frustrations

- Existing solutions can't efficiently process long, complex legal documents.
- Existing tools require manual copying and re-uploading instead of allowing direct in-tool edits, making corrections time-consuming.

> I don't want to just catch AI use - I want to help students understand and improve their writing.

Professor James Carter

50, New York

Occupation: University Professor (English & Literature)

Education: State University of New York (SUNY)

Bio

Experienced English & Literature professor at SUNY with a focus on academic integrity and AI detection in student work. Skilled in AI content verification, digital learning tools, and student engagement. Regularly uses Turnitin, Google Docs, and Canvas LMS to assess originality but seeks a faster, more transparent AI detection tool with LMS integration and clear, student-friendly feedback. Passionate about guiding students in ethical writing while leveraging technology for fair assessments.

Goals

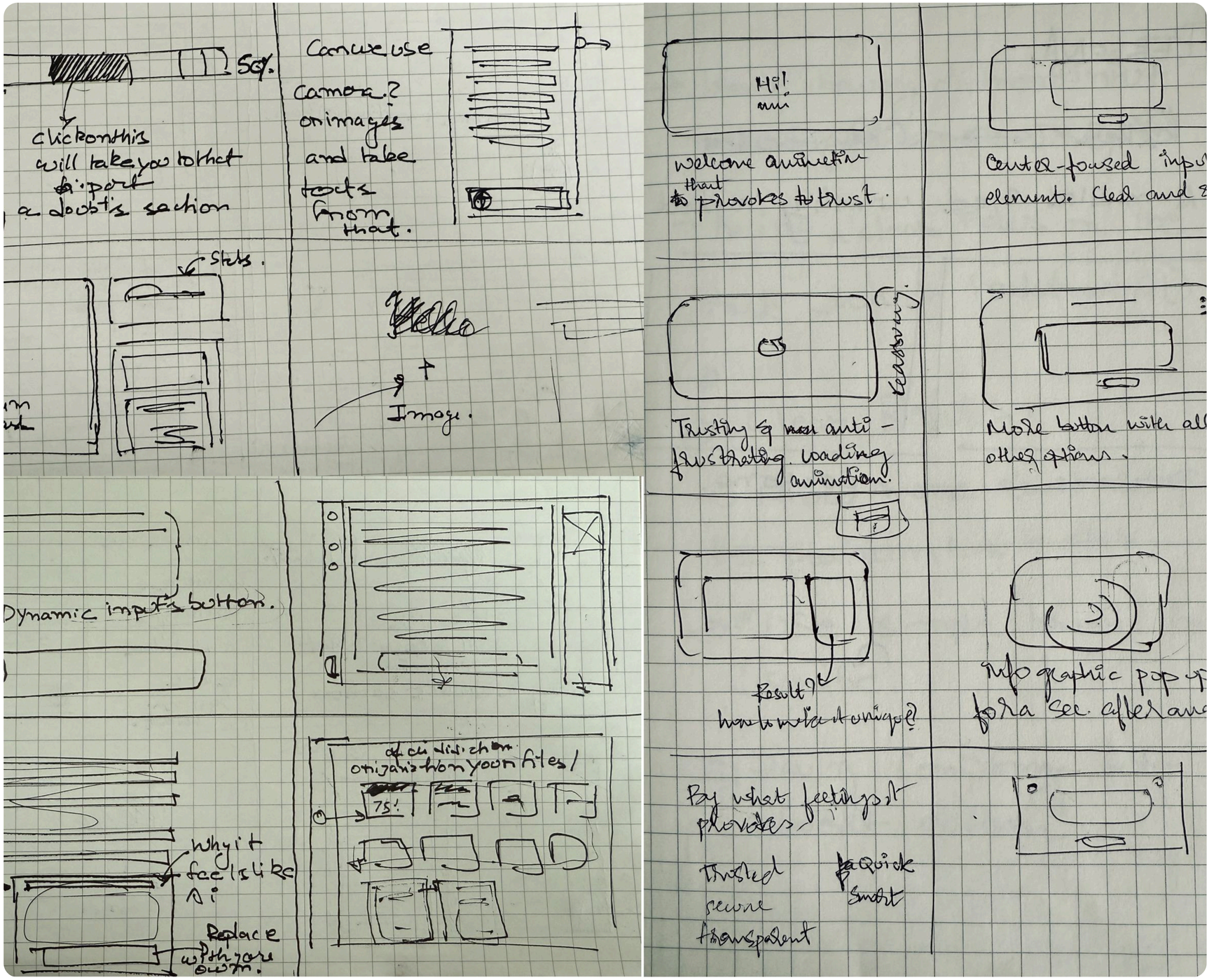
- Quick AI Content Detection: Needs a fast, reliable way to analyze student essays.
- Transparency in AI Score: Wants a clear breakdown of flagged content (not just an overall % score).
- Integration with LMS: Should be able to upload files from Canvas, Google Drive, or Dropbox easily.

TECH PROFICIENCY

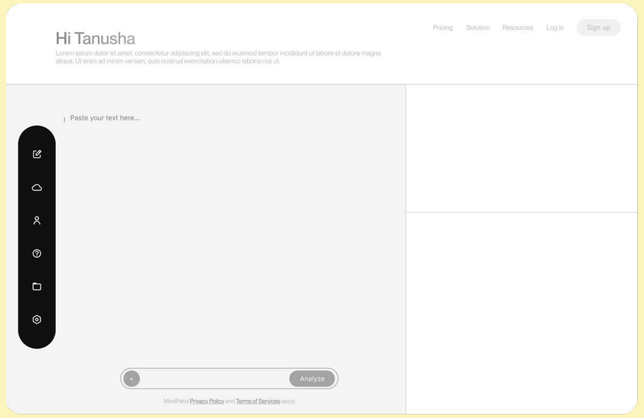
Frustrations

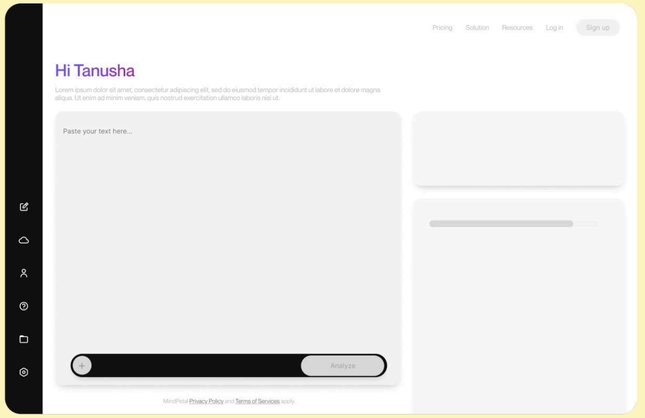
- AI-generated content is becoming harder to detect manually.
  Lack of Transparency: Some tools provide a probability score without explaining where a document is AI-generated.

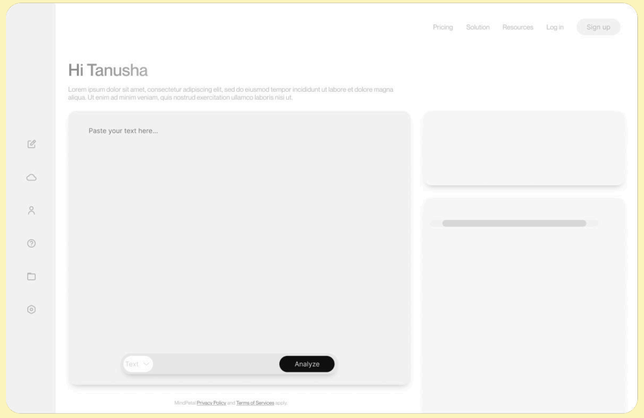
Sketch:
Explored a wide range of ideas through sketching and ideation exercises like Crazy 8s.

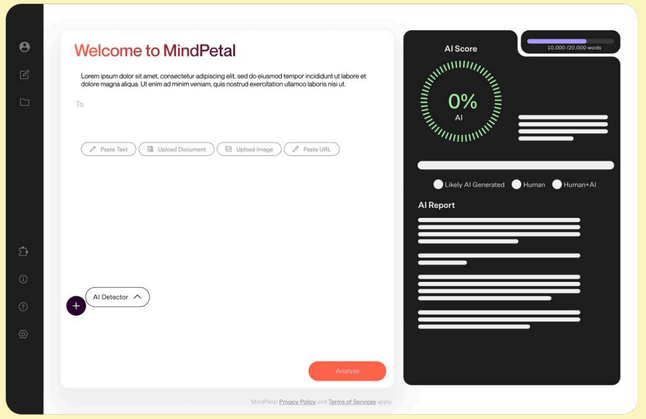


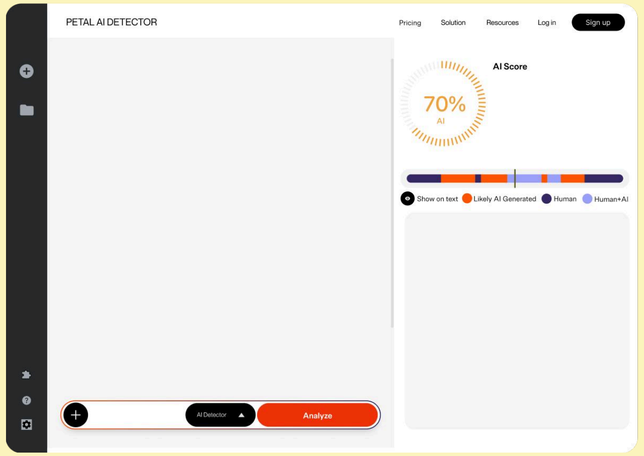
Wireframing: Experimented with numerous layouts and structures, refining designs through multiple iterations to achieve the best user experience.

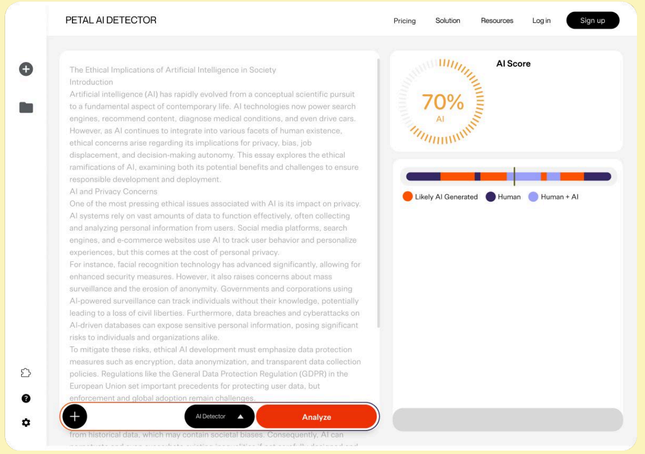












25+

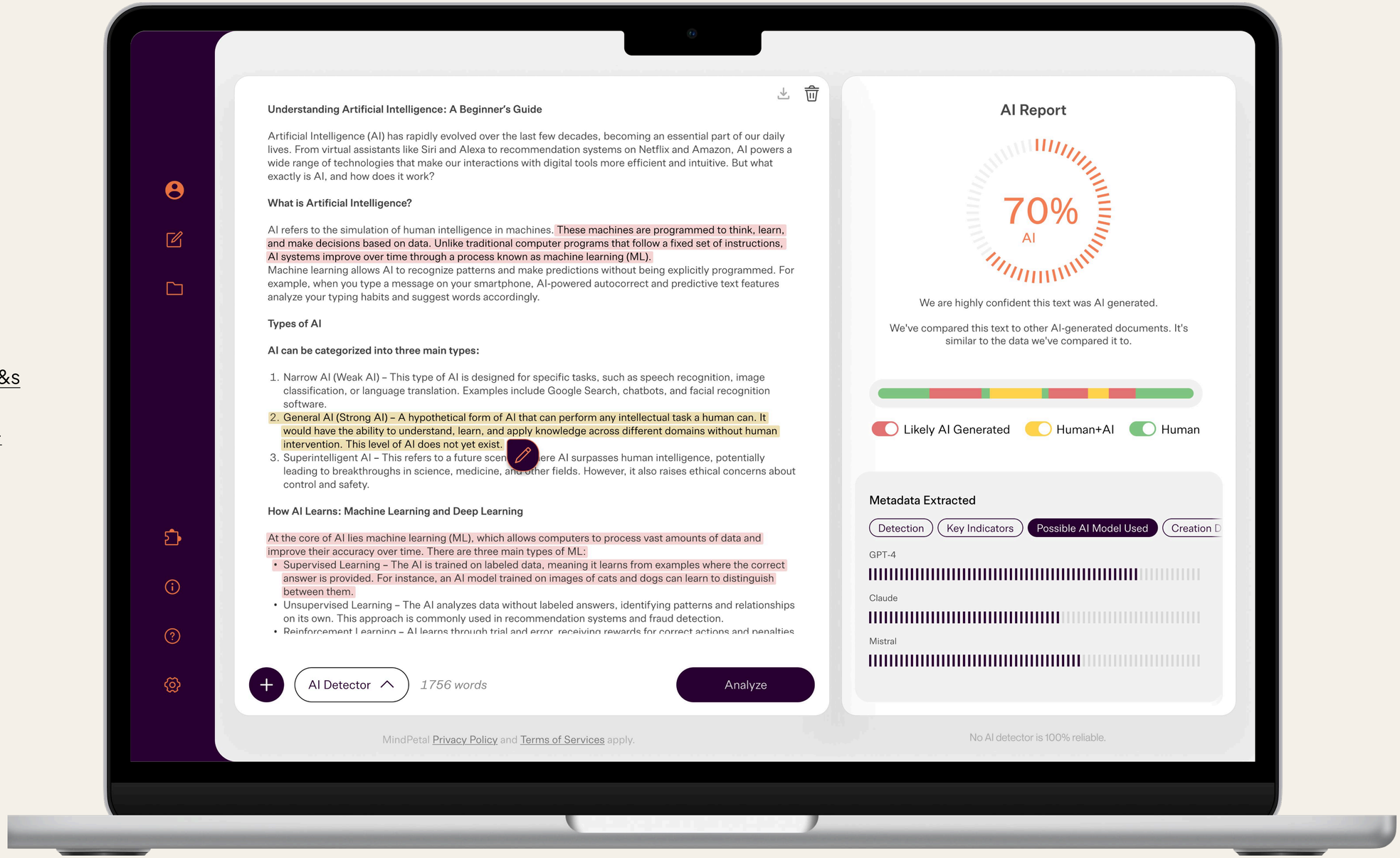
Prototype

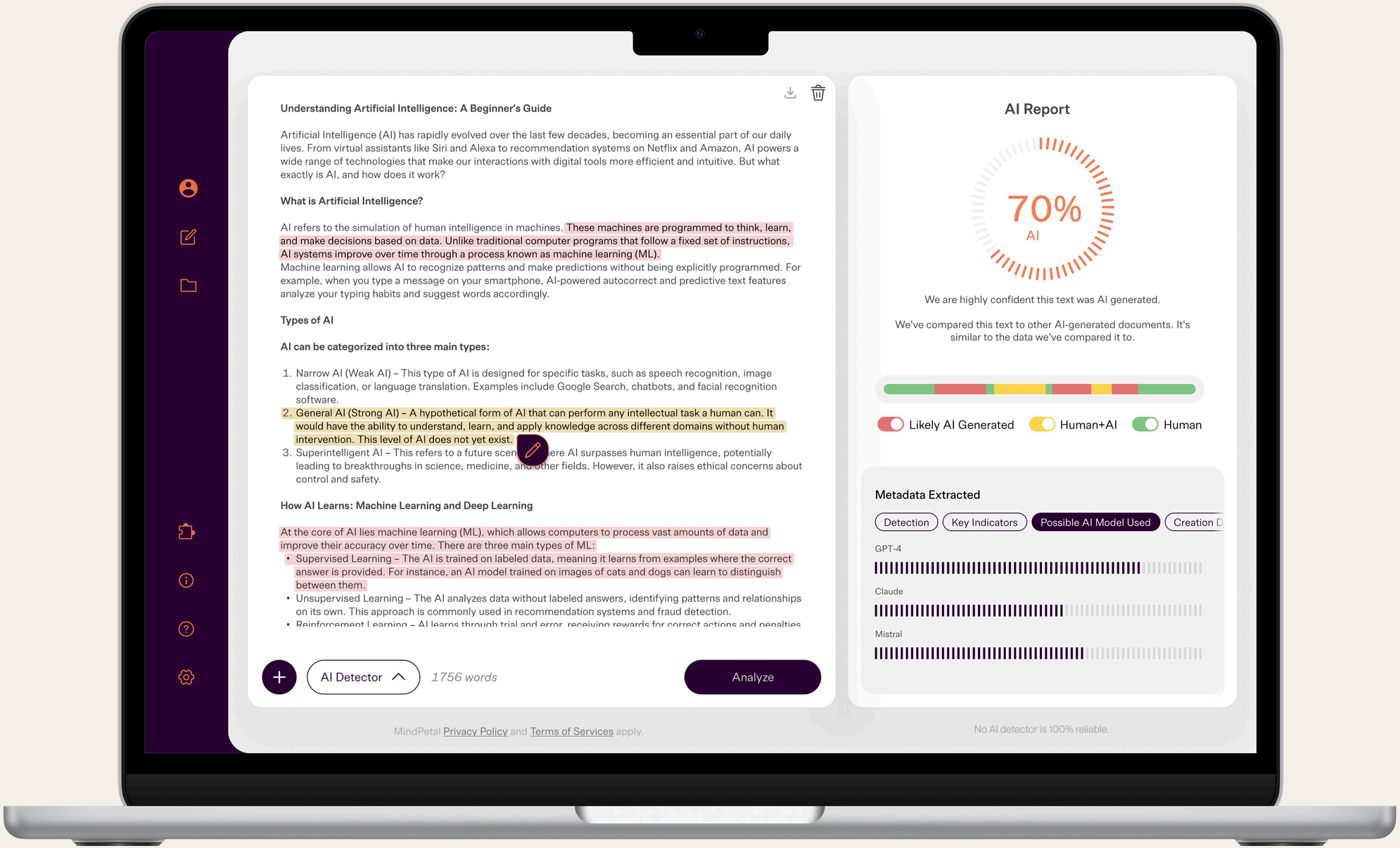
Figma link

https://www.figma.com/proto/BVFe5IbrnONpxlGMR0ccZQ/IC25---AI-Recognition-tool?node-id=3-11003&p=f&t=zcdQ8ZTusqxPzkFI-1&scaling=scale-down&content-scaling=fixed&page-id=0%3A1&starting-point-node-id=3%3A11003

Prototype Video link

https://vimeo.com/1063822561?share=copy#t=0




Understanding Artificial Intelligence: A Beginner's Guide
Artificial Intelligence (AI) has rapidly evolved over the last few decades, becoming an essential part of our daily lives. From virtual assistants like Siri and Alexa to recommendation systems on Netflix and Amazon, AI powers a wide range of technologies that make our interactions with digital tools more efficient and intuitive. But what exactly is AI, and how does it work?
What is Artificial Intelligence?
AI refers to the simulation of human intelligence in machines. These machines are programmed to think, learn, and make decisions based on data. Unlike traditional computer programs that follow a fixed set of instructions, AI systems improve over time through a process known as machine learning (ML).
Machine learning allows AI to recognize patterns and make predictions without being explicitly programmed. For example, when you type a message on your smartphone, AI-powered autocorrect and predictive text features analyze your typing habits and suggest words accordingly.
Types of AI
AI can be categorized into three main types:
1. Narrow AI (Weak AI) – This type of AI is designed for specific tasks, such as speech recognition, image classification, or language translation. Examples include Google Search, chatbots, and facial recognition software.
2. General AI (Strong AI) – A hypothetical form of AI that can perform any intellectual task a human can. It would have the ability to understand, learn, and apply knowledge across different domains without human intervention. This level of AI does not yet exist.
3. Superintelligent AI – This refers to a future scen ere AI surpasses human intelligence, potentially leading to breakthroughs in science, medicine, and other fields. However, it also raises ethical concerns about control and safety.
How AI Learns: Machine Learning and Deep Learning
At the core of AI lies machine learning (ML), which allows computers to process vast amounts of data and improve their accuracy over time. There are three main types of ML:
• Supervised Learning – The AI is trained on labeled data, meaning it learns from examples where the correct answer is provided. For instance, an AI model trained on images of cats and dogs can learn to distinguish between them.
• Unsupervised Learning – The AI analyzes data without labeled answers, identifying patterns and relationships on its own. This approach is commonly used in recommendation systems and fraud detection.
• Reinforcement Learning – AI learns through trial and error, receiving rewards for correct actions and penalties
AI Detector
1756 words
Analyze
AI Report
70%
AI
We are highly confident this text was AI generated.
We've compared this text to other AI-generated documents. It's similar to the data we've compared it to.
Likely AI Generated
Human+AI
Human
Metadata Extracted
Detection
Key Indicators
Possible AI Model Used
Creation D
GPT-4
Claude
Mistral
MindPetal Privacy Policy and Terms of Services apply.
No AI detector is 100% reliable.

Q&A

Thank you!