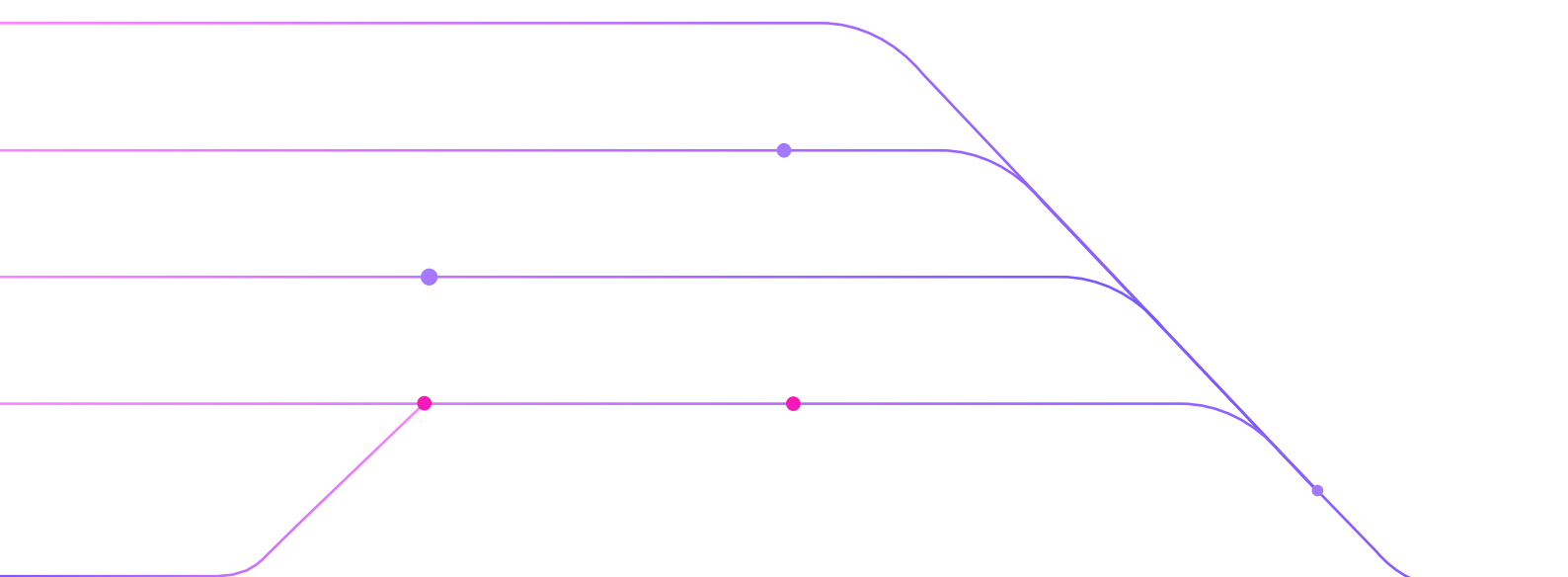


Human Protocol **AI Whitepaper**

contact@hmt.ai

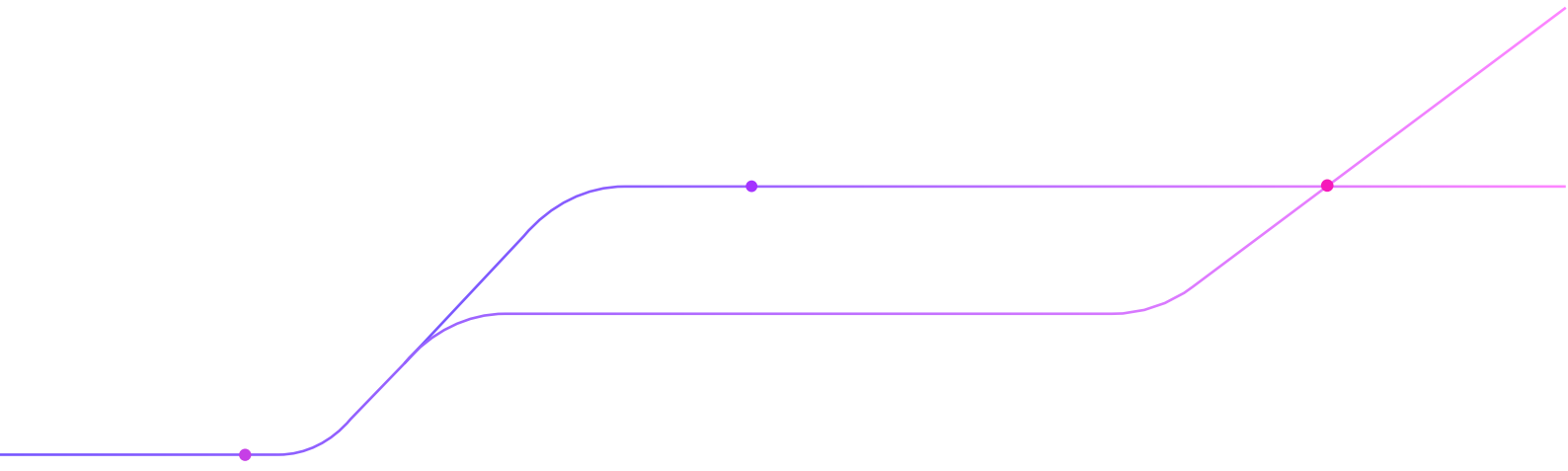
Version 1.0 DRAFT

Contents



03	Introduction
04	Motivating Questions
05	Possible Answers
06	Machine Comprehension
14	The Platform Pipeline
20	Mission
22	Goals

Introduction



The HUMAN Protocol is engineered to allow all kinds of labor to be traded across automated, distributed marketplaces. However, we have focused initial applications of the Protocol on marketplaces where huge volumes of human annotation can be traded in order to facilitate machine learning.

In this Whitepaper, we delve into the difficulties of training and achieving machine comprehension, and explore how an automated HUMAN platform can facilitate the next expansion in AI capabilities: for machines to request the data needed to improve their algorithms.

Motivating Questions

1. If you're building machine learning models, could the software determine that it doesn't know what it doesn't know? Why couldn't the AI ask directly for clarification?
2. When we see an edge case or error, can we ask: "what is the logical fallacy or gap that has led to this error," and use that information to ask massively distributed networks of people for clarification?
3. Can we build a decentralized web-scale marketplace that solves these problems?

Possible Answers

- The software cannot ask for clarification today because it doesn't have a standardized format to request human inference as it is necessary.
- It is possible to detect errors by querying large networks; it is possible to map these errors to underlying data gaps to direct learning. Mapping these differences requires a heterogeneous, large scale pool of individuals to request clarification from – HUMAN Protocol currently has over 100M individuals answering questions each month.
- HUMAN Protocol defines contracts in software that manages both quality control and payments, interacting with humans who answer questions in 274 countries and territories.

Letting machines ask people for the questions they need to improve is the fundamental gateway to the next breakthrough in AI.

Machine Comprehension

The utility of AI products is founded upon their capacity to understand content – what is meant when a car’s left indicator flashes, or the meaning of a spoken word, such as ‘rain’. However, it is the context in which the content exists that gives it meaning – an indicator flashing left, when the car turns right, is an expansion of context; just as what may sound like ‘rain’ is actually ‘reign’ when we’re talking about monarchies.

For AI to be effective, it must possess a sophisticated way of understanding context. Not only does this highlight the pertinence of more data points, it draws attention to the parameters on which AI is created, which defines its understanding of context. HUMAN offers a bottom-up solution; it allows millions of humans to contribute more data to machine learning practitioners, while also assisting in the problem of the limited outlooks of those who train and write the AI algorithms, by providing a system in which the computers themselves can ask for the data they need.

Why is comprehension hard to attain and hard to evaluate?

If we had true comprehension of text, translation between English and Chinese would be reliable and good.

- If you met a man 5 years ago and asked ‘what is your wife’s job,’ how many assumptions would be layered within that one question?
- For each example we can identify, how many examples are we blind to?

The reason we can point out these examples, and understand them, is only because our personal experiences include that context.

- When you ask individuals in Miami or Saudi Arabia, “is this a long sleeve shirt,” you receive different definitions.

The training data, libraries, and ontologies we give to our AIs have been shaped and hard-coded by a narrow sliver of our population and reflect the experiences and biases of the author(s): grad students, data entry specialists, architects etc.

This results in blindspots, “edge cases,” and encoded bias: an indelible

fingerprint of individuals who defined the library or asset. This is especially well observed in healthcare where data indexing performance is a function of not only contextual information, but also of the strength and extent of mappings of synonyms and hierarchies between concepts (ie there are 12,000+ coded concepts, including a specific code 33XD, which is a billable code used to specify a medical diagnosis of sucked into jet engine, subsequent encounter), so when there is an incorrect response it is far more obvious).

- But, in healthcare, disorders and diseases are constantly researched and re-classified as we learn more, and
- individuals or expert panels are not sufficient to define and redefine these relationships.

Concepts are not only mapped and evaluated with tools such as graph analysis, but also through supplying of context through libraries. This has resulted in proprietary “black-box” libraries and methodologies for refining and layering upon the ways in which concepts are linked and mapped – much like an apprentice relationship, this information is frequently passed from scientist to algorithm through added layers of context supplied as conditions.

Context provides meaningful and important information, and in many ways these issues are more obvious in specialized domains because of

the robustness and variety of standards for classification (ie, failures aren't silent, they are explicit in healthcare, and we can learn from them).

Example: "AFP" in healthcare:

- in a paragraph about the kidney and labs, AFP most likely stands for "Alpha FetoProtein," a lab,
- but in a paragraph about symptoms and the face, it is more likely to mean "Atypical Facial Pain."
- Healthcare, as a field, only elucidates these failures because robust coded libraries proliferate and this makes the failures more obvious.

Example: Mr. Huntington has Huntington's Disease and lives on Huntington Street.

- Huntington means three different things, depending on context and these terms may be repeated in a variety of different sentences throughout the document.

This applies not only to text, but also to visual representation learning. It is easy for It is a challenge to label and understand these differences throughout not only the structured elements, but also through unstructured writing

- Proper classification is important for being able to achieve data interoperability and utility:
- Proper classification is essential to higher order functions, such as useful predictive and outcomes modeling in healthcare and other specific domains.
- Attempting to load data into a visualization dashboard such as Snowflake can illustrate these issues rapidly.

Because of the robustness of tools in healthcare and the work on libraries in this arena, failures do not happen silently. Similarly robust tools do not yet exist for many other domains.

In this example, several factors influence performance

1. Identification of document section
2. Identification and refinement of relationships between context
3. frequently the author identifies the edge cases and then creates more robust conceptual maps or contextual classifiers to address that edge case.

Context is not only important to comprehension and translation, but also feeds back into adequate solutions for problems like refining and improving handwriting recognition. Today, our contextual bias is hard coded into the conditions and libraries we supply to our algorithms.

Example: only half of people consider a truck to be a type of car, but most humans consider both trucks and cars to be types of automobiles.

Computers to identify images which contain a man or a dog or both, but hard to identify that an image is of a man petting a dog or playing with a dog.

- Many categories of tasks involve understanding the physical world and interactions in the physical world via embodied agents or proxy learning from humans or robotic bodies i.e. some concepts are beyond linguistic representation.
- Furthermore the concept of cultural context extends to the visual world, consider why airplane stewards and stewardesses will point with two fingers instead of one, because pointing means different things in different regions and can even be offensive for some.
- Understanding of abstract concepts and reasoning in the world without assigning language symbols is an area of interest and research, and the overlap between these domains is useful when it comes to human labeling tasks.

While it is easier to demonstrate this issue in the field of language, it rings true for symbolic systems and the portability of concepts around language. There exists a hidden translation layer between language / visual input and meaning which is broadly encoded in societal context.

Beyond translation/communication, these same issues apply to scene understanding, segmentation, detection, information retrieval, quality evaluation, understanding of human preferencing, tracking, measurement, perception for planning and even real-time decision making.

Multimodal conceptual understanding and transfer learning will be essential to progress in similarity-based learning. Visual + Linguistic (V+L) learning is an area of deep interest because:

- There is a wealth of video-based data available online for learning
- Video or live-experience is a primary mode of learning for humans
- Trends in machine learning surround embodied learning
 - HUMANS learn and our brains have evolved through interactions in the world
 - V + L connects to this in a shortcut way through examples of videos and curriculums of videos that illustrate objects in their natural environments, behaviors etc are useful in associating things in the real world and may be useful for reinforcement learning.
 - Can we detect shortcut learning (through explicable AI) and use that to determine if shortcuts should be valid or invalid?
 - Beyond textual hierarchies, the combination of conceptual relationships and proxy real-world experience is a powerful component which has shaped the evolution of the human mind, therefore we believe it is likely to be a helpful component of curriculums for reinforcement learning.

Everything derives from a fundamental understanding, higher order layers should be pulling from much more robust contextual information:

- If we ‘understand’ that “put one foot in front of the other” means “to take small steps to achieve a goal,” then translation of this concept would occur properly.
- And If we solve these underlying contextual meta-translation issues and solve at this deeper layer, higher order tasks would improve.

Most people have exposure only to one such societal context - there is no one person who can create the standard because they have not lived a million lives in a thousand places. As such, the only way to create a robust meta-translation layer is through massively distributed technology - together we must define the structure of that translation layer and populate it with relational information.

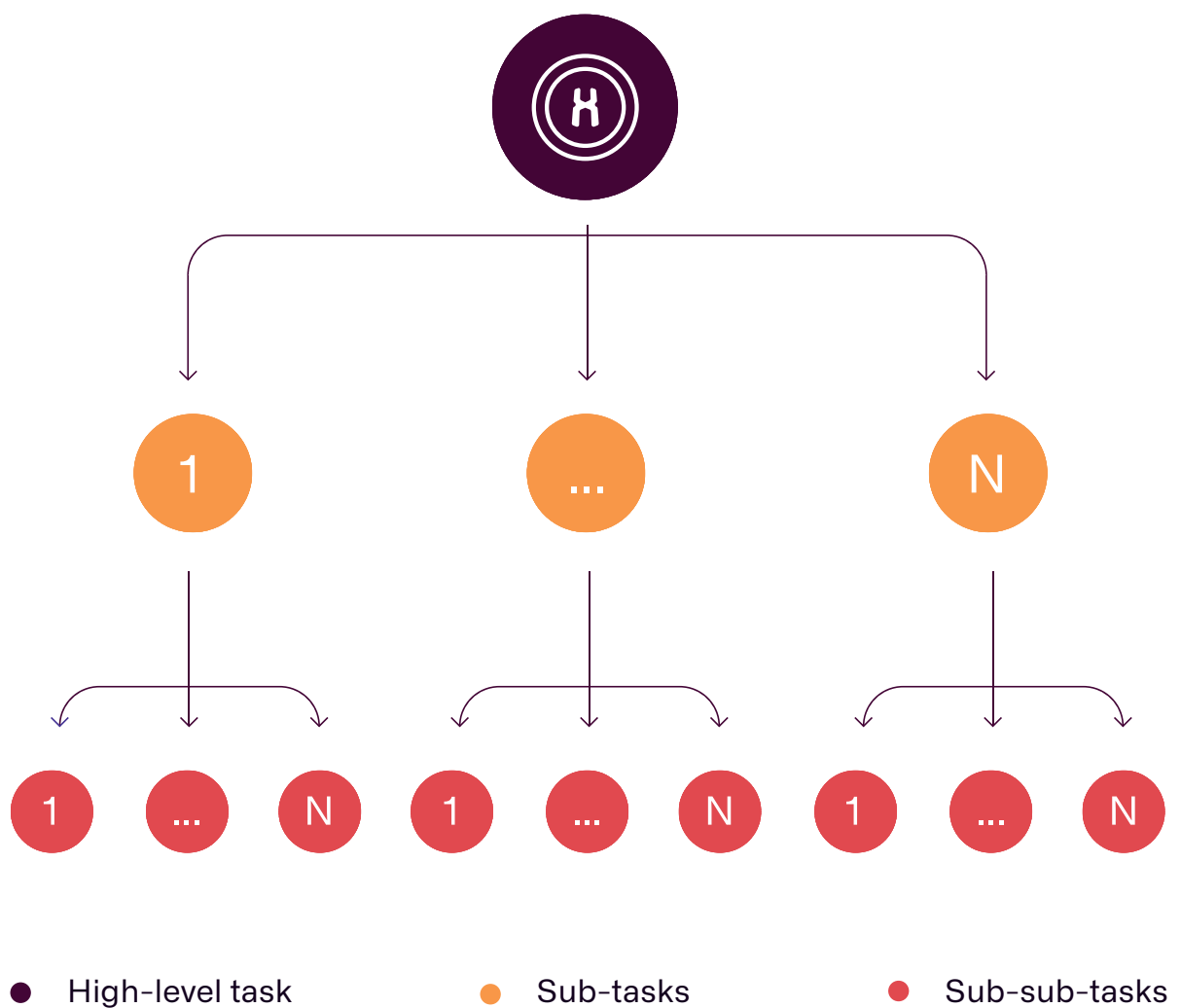
For us to create broader dialogue between AI and HUMANity, in a global context, we must make concepts and education more approachable for a wider group of individuals, prioritizing collaborative efforts around education and pursuing simplicity in our explanations.

The Platform Pipeline

The HUMAN Protocol is a new way to organize and access human labor, with a focus on human-machine interaction for AI.

The HUMAN Protocol allows job types to be contractually defined and broken down into smaller component tasks ("factored cognition"), with funds placed in escrow (paid out only for accuracy above a threshold) — in effect, this enforces guarantees on annotation quality and execution and enables AI to request human insight in real-time.

Factored Cognition



The HUMAN Protocol has:

- Open source community and audited code at <https://github.com/hcaptcha>
- Over 100 million monthly active users completing tasks
- HUMAN platform technology is already in use on ~15% of the internet via its first application, hCaptcha, with daily interactions with many millions of users across tens of millions of websites.
- HUMANS engaged in task completion represent 247 countries and territories, including one human in the South Sandwich Islands (a small inhospitable island of 30 people – so, by percentage, it's still pretty good).

This generates an enormous volume of human annotation capacity for use in machine learning, accessible by API. New services can be launched at web scale with “human quality” inference from day one, with humans in the loop: as models improve in accuracy, continuous QA can keep result quality high.

The Protocol is designed for easy pluggability, allowing you to bring your own interfaces, publish new job types, onboard labor pools, and more, or map your existing workflows onto standardized “job primitives” for the widest universal capacity. Furthermore, instead of building anew, whenever possible we want to aim to integrate with best-in-breed tools within the ecosystem and would love referrals not only in language and visual but in audio and other areas of research as well.

Examples of job primitives include:

- Multimodal similarity recognition between images, text, audio, and video
 - “Does the video contain a dog”
 - “Is the human petting the dog”
 - Matching between domains
 - Given an image can you select the appropriate caption?
 - Given a caption can you select the appropriate image?
 - Sorting tasks
 - Click on the three most similar
 - Find the products that are most similar to the other products
 - Rank the items in terms of similarity

- Validation vs Generation
 - “Draw an bounding box around a dog” vs “Is the bounding box tightly enclosing the dog”
 - “Which summary is more accurate” vs “summarize this text in one sentence”

The underlying technology is now being open sourced, creating an open ecosystem of labor pools of many types, forming a price-transparent global marketplace that makes labor accessible and fungible in a previously unimaginable way.

The HUMAN Protocol is blockchain-based and supports more transactions than Ethereum and more transactions than the sum of transactions for the last 5 years on MakerDAO.

Together, HUMAN and AIs can create not only less brittle, more dynamic conceptual libraries, but also more opportunities for reinforcement learning – for example, determining which translation, or which summary is better – can we make it easier for reinforcement learning with more explicit user-generated rewards?

- Let’s say we have a generative text model, it produces a summary, and then we get a user to rank or select their preferred summary, then we could use those responses as rewards to help generate better summaries.

Mission

With Representation Learning (attaching meaning to words, phrases, visual information), we assume we have effective communication, but often we are misunderstanding one another.

- How many times have you left a meeting, only to later realize you were not on the 'same page' with a colleague?
- Example: "Beverly Hills":
 - Beverly Hills could be properly classified as sensitive/identifying information for redaction from a document (high performance of algorithm for the task), and yet still be misunderstood by the system:
 - Beverly Hills could be either a name or a place dependent on contextual clues

- o In this example, the algorithm could return the correct response (high model performance) and yet, there is not shared ‘understanding’. Until you pursue clarification, and search for silent failures, you cannot be certain you’re talking about the same thing.

And, until you can translate from English to Chinese reliably, we can tell that we have not yet achieved comprehension or effective communication.

In many ways, it appears that humans and machines are still ‘talking past one another,’ and we must make it a priority to ensure that HUMANS and Machines not only speak to one another, but understand one another, as early as possible (and for all of us who are not yet ready for a brain implant). The HUMAN Foundation Team believes this can be achieved by leveraging human inference to train artificial intelligence and machine learning systems.

Goals

Based on an evaluation of complementary strengths we believe there is a rapid path not only to better reading comprehension and global translation, but also being able to summarize and translate reliably, from handwritten notes (which will be important for Augmented Reality applications).

HUMAN Protocol already produces more labeled image data than the whole of Stanford's ImageNet on a daily basis.

Our goals are:

1. Read, summarize and translate from many forms of communication, accurately.
2. Reduce bias & racism in algorithms.
3. Improve communication between humans and AI.

We are an open-source, cloud-first organization aimed at creating shared, public global resources for machine learning. We believe time is of the essence and are community and collaboration-focused.