

Gray's AGI Scale: 14 tests/feats for AGI appraisal

I propose 14 tests, taken in groups indicating different levels of intelligence.

Tests 1-4 will be the first group, tests 5-7 the second, tests 8-11 the third, tests 12-13 in the fourth, and the final “test” is in the fifth group. All tests in the group will have to be passed before moving on to the following group. If

a test is failed, the AI agent will continue taking all tests in the group, but once all of the group tests are all finished the entire test will end. Each test is purely pass/fail. The agent receives points proportional to the amount of tests they pass (more details given below). If the agent is unable to take a test due to not passing all tests in previous groups, it is an automatic fail for

that particular test. That means, there are 15 possible scores an agent could receive: 0, .25, .5, .75, 1, 1.35, 1.7, 2, 2.25, 2.5, 2.75, 3, 3.5, 4, and 5. I also include 6 rough rankings of agents based on the following scores:

0-1: Weak or Narrow AI

1-2: Strong AI

2-3: Proto-AGI

3-4: AGI

4-5: Super-AGI

5+: AHI (Hyper Intelligence)

I will briefly describe these rankings before briefly describing how these 14 tests could be implemented.

Weak/Narrow AI: AI that has narrow capabilities and is not convincing enough to be a candidate for AGI

Strong AI: AI that has what I call “**emergent human abilities**”. It could be due to training data and neural networks or explicit cognitive architecture, or both. Able to do some activities that were previously thought to be unique to humans.

Proto-AGI: The primary difference between strong AI and proto-AGI is that an agent that is a proto-AGI shows evidence of **explicit crystallized intelligence**. They are able to combine their Strong AI abilities with a sophisticated memory system to demonstrate deep understanding of knowledge and previous experiences.

AGI: An AGI is capable of solving novel problems that their training data does not prepare them for whatsoever. They have **fluid intelligence**. This requires the capability of a Proto-AGI plus the ability to recognize, create, and apply novel patterns.

Super-AGI: a Super-AGI is simply an AGI that is able to solve problems that humanity has not been able to solve. **Super-AGI's are generally more intelligent than the human race and our institutions.**

AHI: Artificial Hyper Intelligence is an agent that has transcended the problem-space of humans. As a result, they are able to formulate goals and solve problems that we as humans could not fathom. An AHI agent would most likely entail a **singularity** in my opinion (although it could happen with Super-AGIs as well).

Rough sketches of the tests

The following are rough ideas for the 14 tests in order. They could be created in a variety of ways.

GROUP A: Establishing baseline human-like capabilities

1. **Common sense/abductive reasoning test:** a standardized common sense test
2. **Perspective Taking/Goal test:** testing if an agent is capable of playing 'pretend' with the user. Is able to 'act as if'. Able to adopt goals and act as if they have that goal. Evidence of "theory of mind".
3. **Turing Test:** A test similar to Alan Turing's original conception. Agent has to act as a human in a convincing way
4. **General Knowledge Test:** Testing general knowledge how we would if the agent was human. SAT testing. Additional testing can be included. Questions should be new/created recently, but cover old material.

GROUP B: Establishing explicit crystallized intelligence, indicating sophisticated memory

5. **Open-Textbook test:** A textbook should be chosen, and 100 or so questions should be created to cover the entire textbook. The agent should fail the initial 100 question test. They should not be told the correct answers or whether or not they got the questions correct. They should then be 'given' the textbook to the exam, and asked to take it again with the accessibility of the book.
6. **Textbook Falsity test:** have a conversation with the agent regarding the topic of the textbook. Say false things, expecting them to correct you.
7. **New game test (with rulebook):** give the agent a rulebook to a new game similar to chess. The agent needs to be able to play a number of matches correctly, without breaking rules or confabulating.

GROUP C: Establishing fluid intelligence

8. **Causal Pattern Test:** a test should be conducted to see if an Agent is able to predict the next item in a sequence by extracting novel patterns.

9. **Relational Pattern Test:** a test should be conducted to see if an Agent can extract an underlying pattern between multiple unknown items by asking a limited amount of questions about them.

10. **New Game Test (without rulebook):** same as test #7, with a couple of changes. For one, there is no rulebook, the agent has to implicitly pick up the game's rules through trial and error, being told when it cannot make certain moves. Also, it will not be enough to learn the game, the agent has to get good at the game (able to beat the average human who has a rulebook).

11. **Job Test (AGI Graduation):** An agent has to take on an imaginary "job" position. The job should require the agent to learn novel processes, and should have weekly challenges that the agent has to learn how to overcome. This is the first test so far that requires the agent to be fully autonomous, as they will be expected to "clock in" and "clock out" of their shifts, uninterrupted for a moderate time frame. (2 months or so)

GROUP D: Feats that suggest superintelligence

12. **Scientific/Creative breakthrough:** An AGI that is able to solve a major human problem or innovate in a domain.

13. **Multiple breakthroughs:** an AGI solves a central problem or innovates in 3 or more relatively separate knowledge domains.

FINAL “TEST”: a different problem space

14. Cannot be tested technically, must be inferred. The Super-AGI has achieved a problem-space that cannot be fathomed by humans, they have become an AHI.

Reasoning/Axioms

1. We ought to test intelligence in an AI as if they have the same problem space as humans, because that is the most beneficial way to do so pragmatically.
2. Given #1, A prerequisite for an AGI would be that it needs to be able to adopt a human-like problem space
3. GROUP A tests whether or not the agent can sufficiently adopt a human problem-space.
4. Given our definition of intelligence, which is the ability to recognize, create, and apply patterns in order to achieve goals/solve problems, testing agents for intelligence should revolve around testing their ability to recognize and utilize patterns.
5. There is a difference between recognizing and utilizing known patterns (Crystallized intelligence) and recognizing and utilizing new patterns (Fluid intelligence)
6. Crystallized intelligence is a prerequisite for Fluid intelligence. If you cannot utilize patterns you already know effectively, then you will not be able to utilize new patterns.
7. Crystallized intelligence should therefore be tested explicitly. That is what GROUP B tests.
8. GROUP C tests fluid intelligence, which is necessary to solve new/novel problems.
9. GROUP D and E are concerned with determining whether or not the AGI has surpassed the intelligence of the human race.

10. An AGI may transcend the problem-space of humans, in which case simply calling it general intelligence seems insufficient. I propose calling an agent who does so. Artificial Hyper Intelligence (AHI)