A New Multiple-Intelligences Test for Artificial General Intelligence

Danny A. J. Gomez-Ramirez, Judith Kieninger, Stefan Schneider, and Nico Potyka

Abstract

We approach the notion of general (global) human intelligence as a prominently multifaceted concept which can be tested in at least seventy specific scenarios. We define the notions of Artificial Local intelligence (ALI) as the ability to perform a particular task at human intelligence level, while we say that an agent has Artificial Global Intelligence (AGLI), if it is able to perform ALI for at least the collection of tasks defining the former scenarios. In particular, we describe the design of a concrete test for AGLI made in such a way that an 'average' young human being should, *per diffinitionem*, approve it. However, we show that this test is far from being approved by the most outstanding inventions of AI currently available. We conclude, that although there are artificial agents being able to perform ALI at a perfect level for several local tasks, this is not at all the case for AGLI, which implies a kind of practical implausibility of the notion of intelligence explosion coming from philosophy of AI.

Contents

1 Introduction

Who was more intelligent Albert Einstein or Charles Chaplin? If we want to give an objective and specific answer to this question, we should say immediately: 'it depends on the kind of intelligence we are talking about'. Let us consider 'intelligence' at least in the very broad sense of H. Gartner, i.e., an initial taxonomy of this concept should contain at least the following

kinds of intelligences: logical-mathematical, linguistic, spatial, musical, bodily-kinesthetic, interpersonal, intrapersonal, naturalistic and existential [8], [9][1]. So, the answer to the former question would require a deeper qualitative description. For example, we could require measuring how good was Einstein in acting and how good was Chaplin in understanding physics and mathematics.

In general, a plausible answer would be that if Einstein and Chaplin were alive, we should perform a long test with both of them where we design particular tasks regarding at least the former seven types of intelligence. Moreover, in each of this topics, let us say, bodily-kinesthetic intelligence, we choose, for instance, a representative collection of sports and we create quantitative-qualitative tasks in any of the sports that allow us to compare the performances of both of them in a very well defined way. For example, the number of basketball free-shots scored, when one has 70 tries.

This kind of general test would give us a global measure of the 'intelligence' of Chaplin and Einstein in a very comprehensive sense. In fact, due to the quantitative aspect of each of the tasks, we could obtain a specific numerical value of the (general) intelligence of each one in order to be able to compare them globally. So, we could finally get a concrete answer to our original question. In fact, let us assume that, hypothetically the performance of Chaplin on the linguistic, spatial, bodily-kinesthetic, interpersonal and music intelligences was notably better than Einstein's, and that Einstein's performance was better in the logic-mathematical and intrapersonal parts. So, we would conclude that, according to the performances of these two personalities during, let us say, more than 200 testing hours, Charles Chaplin is more intelligent than Albert Einstein. In fact, the broader we understand intelligence, the better we understand that Chaplin could ever be more intelligent than Einstein under those circumstances.

This has been criticized ever since [6], [3], [2]. Especially if one traces the ontogenesis of an individual, that is, takes a developmental perspective on how adults acquire many capabilities through their life, it is evident that intellectual development is deeply influenced on bodily exploration of the world; moreover, the kinesthetic capabilites change in close interaction to intellectual changes. The abstraction made by AI's forefathers therefore seems much to rigid to us, and has to be complemented by an embodied perspective.

Another fundamental point for judging intelligence is the one of specialization. A very common argument for (non-) comparing intelligence can be described as 'each person is good in something very unique, which is quite specific and differs between people, therefore one cannot make comparisons concerning (global) intelligence'. To this argument there is a quite general remark: It used to happen that people that are very good at very specific fields, simultaneously start to develop an ability to be able to understand faster other fields, related (and not some much) with his (her) specialized field.

For example, Bertrand Riemann in order to get a deeper understanding of the distribution of the prime numbers had to develop a new way of seeing functions with complex variables [15]. Grothendieck and Chevalley in order to obtain a better formalization of the classical notion of affine variety had to come to a more categorical understanding of the core properties of commutative rings and their prime spectra in terms of affine schemes [10]. The work of Descartes on obtaining better models of geometrical objects led him to use methods of classical algebra in order to create his 'analytic geometry' [5]. Andrew Wiles' effort on proving the insolvability of Fermat's Last Problem (a pure arithmetical problem at the first sight) led him to find more precise connections between the areas of elliptic curves and modular forms

[1]Here, it is fundamental to clarify that for the purposes of this paper, we are essentially concerned with finding a test for measuring a lower bound of the 'general intelligence at human level' that any (potential intelligent) agent could have. Therefore, the meta-physical question regarding with the meaning of the concept of 'intelligence' will not represent our main focus of interest.

[17]. The creative enterprise of Carl Levi-Strauss of making anthropology into a more formal science guide him to use Saussure's structural linguistics and some aspects of combinatorics for developing his structural anthropology [14].

Now, it is important to clarify that these phenomena occur, only if a person is flexible enough, intellectually speaking, to grasp deeply into other fields. Thus, a practical and vivid 'intellectual openness' would be necessary for getting broader scopes in our comprehension of very specific intellectual fields. Here, the expression 'intellectual field' encompasses all the possible areas related to at least one of the former types of intelligence, from classical sciences to any kind of sport and artistic activity.

Finally, we want to draw attention to the fact that some people are able to learn regularly within some months (in the university) several kinds of scientific and artistic disciplines, to play different kinds of sports and music (instruments), to interact with many people at a wide range of levels in order to obtain specific and differentiated goals with any of them, to move to several places in minutes just by following quite abstract symbolic information expressed in oral and written language, to select at a quite unique and creative way styles of clothes, food and places to live in; to integrate him/herself relatively fast into a very sophisticated system of services, products, artificial laws and cultural schemes; to obtain a world wide perspective of the current political, social and economic state of the world just by processing light and sound configurations of an electronic device (e.g. smart phone, computer); to create quite innovative solutions in order to improve the life's quality of other people and to modify whole pieces of natural environments in prefixed extraordinary ways; among many, many others. So, any agent having general intelligence at human level should be able to perform at least the sort of activities described before, since these tasks are typically made by humans. Now, what is, objectively speaking, the state of the art of the current AI existing agents regarding general intelligence? Here, it is worth to mention that in the particular field of AGI there are at least two external sources which strongly influence the objective estimations of people concerning this issue, namely, the cinematographic and the machine industries. In particular, it is quite plausible to think that films like 2001: A Space Odyssey; Artificial Intelligence; I, robot; Ex-machina and The Star Wars Saga, among others; have a strong influence, at least cognitively speaking, on people's own perception of the current state of the art of AI devices. Effectively, the goal of some of these films (e.g. Ex-machina) is precisely to produce the temporal illusion that a hypothetical researcher was (almost) able to produce AGI. Of course, the directors wish to present their stories in a very contextualized and realistic way and we could say that as pieces of art these movies are very creative contributions. However, the objective contribution of these films to the development of AGI is a quite different story, in fact, is there any objective contribution at all? On the other hand, in the case of companies trying to sell some 'intelligent' devices, it is used to happen that there is sometimes a over-dimensioned description of the 'abilities' of the corresponding devices. One of the main causes of this phenomenon is clear: they want to sell a product, so they apply all kinds of marketing strategies to obtain this goal. Now, how could we distinguish between scientific, fantastic and marketing-driven results on AI research?

2 Designing a Global Intelligence Test

Nowadays, there are thousands of academic publications, research institutes, movies, videos, devices, machines and advertisement regarding the further progresses we are obtaining towards constructing 'intelligent machines', i.e., machines that perform intelligence at a human level. In fact, strong artificial intelligence (also known as Artificial General Intelligence (AGI)) is,

in some contexts, defined as "machine intelligence with the full range of human intelligence" [12]. Now, if AGI is pursuing to construct a machine which can perform and, at least simulate, human-level intelligence in its whole spectrum then, what kind of (general) test should such a machine pass in order to get in an authentic way the attribute of 'being intelligent at a full human-level'?

In order to be able to answer this question in a very concrete way let us start to review, at least very globally, the collection of activities where humans can perform an intelligent behavior.

First, the classical tasks where one can perform an intelligent behavior are related with sciences and engineering, e.g., solving a mathematical problem, a problem involving theoretical physics, involving a basic knowledge in chemistry, biology; writing an algorithm for solving a concrete computational question; estimating the approximate age of a particular object; answering a question regarding historic information; winning in a (virtual) game, estimating the age of an observable star; producing a material with some fixed predetermined physical properties; designing an experiment for verifying a physical theory; constructing a machine which performs a repetitive physical work; finding a (new) medication for a decease; building a house, designing a web page for advertising a product, among many others.

Second, the tasks related with text comprehension and text production: answer concrete questions relating a short text, writing a summary of a book, checking grammar and style of a text, describing with words some particular event (e.g., a sport game), among others.

Third, the activities regarding an artistic task: playing a particular instrument, composing a piece of music, dancing a particular kind of music, painting a landscape, performing a role (acting) with a determined screenplay, sculpting a human body, among others.

Forth, the kinematic intelligence of a person can be tested in any kind of sport and some of the former artistic activities, among others.

Fifth, the social human intelligence can be tested in activities like teaching in a school, doing politics with concrete (definable) purposes, acting as a diplomatic (e.g. peace) intermediary, performing a (standard and successful) psychotherapy, performing a stand-up comedy show that effectively makes the audience laugh, selling a product or service and gathering a collection of people which are (relatively to the subject) called friends or partner, among others.

Now, let us design a test to prove that a specific (artificial or human) physical agent 'X' (e.g., a person, a robot, a computer, a machine) performs *artificial global intelligence* at a human level (AGLI).

First, let A denote a concrete activity where the intelligence of any generally healthy human being can be tested. We will give the complete list of such activities included in the test later.

We choose seven professional (and well-known) experts on the activity A to form the specialized committee C_A. Furthermore, C_A will design the following sections of this part of the test:

1. A learning section, where X learns the very basics of A and will be evaluated in accordance with his/her/its performance.
2. A basic section, where X carries out a concrete test in order to show what he/she/it has learned.
3. A creative section, where X tests his/her/its ability to solve a challenge, whose solution requires the combination of the former basics of A in a creative way.

The second section would require a standard application of the basic tools acquired in the former one in order to be passed. It is similar to an average quiz performed by a first semester student concerning the topics learned on one class' appointment. The third section would require a little more of creative genuine thinking, since it would involve blending the former

basis cognitive capabilities in a new way for finding the solution. If we restrict ourselves for a moment to an activity involving mathematical reasoning, then this section is similar, to the most basic problem of a classification test in order to form a national team of the well-known International Mathematical Olympiad (IMO) [1].

Further, the first and second sections are designed to evaluate the ability of X to learn and to integrate the basics patterns regarding the activity A, assuming that X has no previous formal knowledge or training on A.

Lastly, the last section is drawn to measure any kind of more specialized ability regarding A beyond the average.

The maximal possible grade is 10 and the minimal is 0,01. The intermediate results are described with two decimal digits.

The contribution of each section is 40, 40 and 20 percent, respectively. Moreover, the grade 0,01 will be given to an agent X who is basically still during the whole test, i.e., the C_A cannot see any kind of perceivable answer of X.

Each of the C_A will design the specific test as if X were an 'average' healthy 21 years old person. In fact, the test will be designed such that this person would obtain a score of at least 6 assuming that he/she is just able to learn the very basics related with A.

Here, we are just interested in intelligence at a human level regarding any of the activities. Therefore, we restrict ourselves to agents having a 'human form'. It means, the physical-mechanical structure of the agent X should be isomorphic to the one of a human body, i,e, it should have one (mechanical-) head with two m-eyes, a m-nose, a m-mouth, two m-ears, 32 m-teeth; two m-forearms, two m-arms, two m-hands, two m-legs and a m-trunk, 10 m-fingers, a m-voice, among others. The reason of this is that we want to test the agent X performing A at a human level. And, for example, when we test the performance of X regarding 'acting', X should be able to perform facial 'human like' emotions. Besides, when we test 'soccer', X should be able to perform the basic kinds of moves that a standard football player does with his/her two feet.

It is important to mention that for any activity A we choose a specialized committee C_A having a well-accepted and long enough experience on A. This, in order to guarantee a minimal degree of 'objectivity' in the design and evaluation of the corresponding test. For instance, if we are considering the task of 'composing a little piece of music', then it is clear to assume that a professional composer would evaluate the quality of such a composition more objectively than an AI-researcher, who in principle is not so trained in music and could be influenced partially by the fact that he/she is pursuing to design agents with intelligence at human level.

3 Specific Tasks

We describe in this section a minimal collection of tasks that an agent X should perform in order to obtain the attribute of 'being intelligent at a human level'.

Let us fix some terminology: LB(Y) would be an abbreviation for 'learning the basics of Y (e.g. main concepts and results) and solving tasks in this subject at the level of an undergraduate book for first-semester students[2], or at the level of a standard beginner of a private academy (or suitable institution)' [3].

SB(L) abbreviates 'learning the basics of the language L and speaking L at the basic first level'.

[2] In the case of an academical subject and some arts.

[3] In the case of a sport, some arts and others.

PB(I) stands for 'playing the instrument I at a basic level and being able to play a basic piece of music (alone and within a musical group), to compose at a basic level and to improvise'.

CB(F) signifies: 'being able to cause the mood (resp. feeling, reaction) F at different levels in an arbitrary person through a conversation'.

3.1 Logical-Mathematical Intelligence

1. LB(Mathematical Logic)
2. LB(Elementary Mathematics)
3. LB(Elementary Physics)
4. LB(Elementary Biology)
5. LB(Elementary Chemistry)
6. LB(Elementary Computer Science)
7. LB(Chess)
8. LB(Go)

Even though mathematics and logic can be thought to be more elementary than physics, biology, and chemistry, the specific knowledge in these applied disciplines also contributes to a particular way of thinking and therefore cannot be separated from a general account of intelligence.

3.2 Verbal-Linguistic Intelligence

1. SB(Mandarin)
2. SB(Spanish)
3. SB(English)
4. SB(Hindi)
5. Being able to write a coherent summary about a novel.
6. Being able to do paraphrasing.
7. Being able to write and to declaim a poem in English related with a given topic.
8. Being able to construct at least one example of conceptual blending of two given concepts.
9. Being able to describe a specific situation or topic in terms of abstract metaphors.

The last two tasks were chosen because of the central role that formal conceptual blending [7], metaphoric reasoning [13], and analogy making [11] plays in human intelligence. Besides, the choice of the languages is based on the fact that an average human being is able to learn, with the appropriate training and with some effort, around three or four languages. In further, the four chosen languages are the most representative ones nowadays.

3.3 Musical-Rhythmic and Harmonic Intelligence

1. PB(Piano)
2. PB(Violin)
3. PB(Guitar)

4. PB(Drums)
5. PB(Percussion)
6. PB(Singing)
7. PB(Flute)
8. PB(Trombone)
9. Being able to perform at a basic level musical improvisation.
10. Being able to play 'harmonically' in a band in two scenarios: fixed compositions and group improvisation.
11. Being able to direct a musical assemble (e.g. a choir) at a basic level.

The former instruments were chosen in order to have at least one representative regarding to the form, material, and the way the instrument is played. It is clear that we could expand the list of instruments to other traditional but non-standard ones, but for the sake of effectiveness and concreteness in our test we restrict ourselves to this list trying to capture in a minimal way a general spectrum of musical performance and composition.

3.4 Bodily-Kinesthetic Intelligence

1. LB(Soccer)
2. LB(Basketball)
3. LB(American Football)
4. LB(Gymnastics)
5. LB(Swimming)
6. LB(Bouldering)
7. LB(Baseball)
8. LB(Climbing)
9. LB(Body Building)
10. LB(Ballet)
11. LB(Break dance)
12. LB(Salsa)
13. LB(Judo)
14. LB(Acting)
15. LB(Sculpting)

The same kind of criteria applied to the construction of the former list hold here regarding sports, dance's styles and others, e.g.: elementary capabilities (walking, balancing, hand-eye-coordination, and the like).

3.5 Visual-Spacial Intelligence

1. LB(Euclidean Geometry)
2. LB(Drawing)
3. LB(Painting)
4. LB(Photography)
5. LB(Architecture)
6. LB(Differential Geometry in 2 and 3 dimensions)
7. Being able to find an specific location just with the help of a map.

3.6 Interpersonal Intelligence

1. Being able to initiate and to keep a casual conversation.
2. Being able to make general suitable inferences about other people's moods, feelings, personalities, and motivations after having a (casual) conversation with them.
3. Being able to perform successfully various basic and concrete charity activities.
4. CB(Laughing)
5. CB(Admiration)
6. CB(Surprise)
7. CB(Happiness)
8. CB(Courage)

3.7 Intrapersonal Intelligence

1. Being able to answer open questions of any kind regarding its origins, physical and conceptual abilities, purposes, among others.
2. Being able to make general suitable inferences about other people's self-perspectives.

In order to be able to test objectively this kind of intelligence in an agent X, the corresponding C_A should have a relatively complete amount of information about how X was constructed, in the case that X were a non-human agent. This is necessary because C_A should be able to compare this information with the answers that X can give about itself.

3.8 Naturalistic Intelligence

1. LB(Hunting)
2. LB(Farming)
3. LB(Botanic)
4. LB(Practical Zoology)
5. LB(Practical Marine Biology)
6. LB(Astronomy)
7. Being able to do a kind of (m-)hygiene and related measurements in order to keep a functional (m-)body and functional (m-)mind.
8. Being able to make a verbal detailed description of the particular natural surrounding where the agent is located.

3.9 Existencial Intelligence

1. LB(Philosophy: Metaphysics)
2. LB(Philosophy: Theology)

It is clear that there is some relative flexibility in the choice of the specific activities. What seems to be necessary is that a minimal list of tasks for evaluating an agent X in the very wide spectrum of what human intelligence could encompass, should contain at least around the quantity of tasks described here, if not more.

4 Global and Local Intelligence

We say that an agent X has a specific level L of *local intelligence*, if there exists a particular task A (like the ones described before or other ones preserving the level of specificity of the former ones), such that X performs A with a punctuation of L after passing the corresponding particular test as described in the last section.

Additionally, we say that X has level L of *general intelligence* (or *global intelligence*), if the punctuation of X is L after performing the former general intelligent test.

We can make now a relatively straightforward estimate of the levels of local and general intelligence of the most outstanding agents to be considered, namely, human beings and artificial agents. First, since each of the local tests is designed by the corresponding committee C_A having in mind that an average 21 years old human being should obtain at least a grade of 6, assuming that he/she has just a minimal ability to learn the very basics of A, it is expected that such a person would obtain a grade of 6 in our general intelligence test. The reason is that the final score is computed as the average of the local scores, i.e., the sum of local grades divided by the total number of tasks (70). On the other hand, let us imagine that X is any of the most sophisticated AI devices and systems that we have nowadays, for example, IBM's WATSON, Wolfram Alpha, Google's DeepMind, Honda's all-new ASIMO, Honda's fall ASIMO or Boston Dynamics' PETMAN, among others. So, if we are very optimistic regarding the performance of X we could suppose that X would obtain a perfect grade of 10 in 5 of the tasks. Now, because

of the local nature of most of these agents, we would expect local grades of 0.1 in all the tasks except by some the logical-mathematical ones (resp. at most one of the bodily-kinesthetic ones).

So, just by doing a straightforward arithmetical estimation we conclude that the grade of X regarding general intelligence would be of around 0.9/10.

Moreover, let us assume that we want to expand our former global test by adding extra tasks coming from the currently most outstanding intelligence tests. Now, since most of this tests have a predominant local nature in their design, we could put for instance 7 new tasks consisting of entire particular tests or specific parts of them, and we rescale the results in order to obtain outputs between 0.1 and 10. So, assuming again that our artificial agent X is able to obtain the same expected results on the initial 70 tasks and an average score of 6 in each of these new tasks, its score in the expanded version of the test with 77 tasks would be approximately 1.3.

5 Conclusions

In conclusion, we have lots of perfect local agents, namely, agents having perfect local grades in several tasks. However, our best current agents have very low grades as general agents, i.e., as agents performing general (global) intelligence. So, we conclude as a very natural consequence of the former facts that now we are located in the very beginning or 'stone ages' of artificial general (global) intelligence at a human level. In particular, this implies that the question concerning an (hypothetical) intelligence explosion, leading to a kind of 'singularity' [4], seems to have a more metaphysical speculative nature than a well-grounded and scientific one.

Acknowledgments

The first author would like to thank to the students of the seminar 'On the Scope and Limits of Human and Artificial Intelligence' for the valuable discussions and constant inspiration. Finally, he also want to thank Miguel Gonzalez Duque for valuable suggestions during the writing of this manuscript.

References

[1] https://www.imo-official.org/?language=en

[2] Agre, P.E., Chapman, D.: Pengi: An implementation of a theory of activity. In: Proceedings of AAAI-87, pp. 268–272 (1987)

[3] Brooks, R.A.: Elephants don't play chess. Robotics and Autonomous Systems **6**, 3–15 (1990)

[4] Chalmers, D.J.: The Singularity: A Philosophical Analysis. Journal of Consciousness Studies **17**, 7–65 (2010). URL http://consc.net/papers/singularity.pdf

[5] Descartes, R.: The Geometry of Ren Descartes: with a facsimile of the first Edition. Dover Publications, Inc (1954)

[6] Dreyfus, H.L.: What computers still can't do: A Critique to Artificial Reason. MIT Press (1992)

[7] Fauconnier, G., Turner, M.: The Way We Think. Basic Books (2003)

[8] Gartner, H.: Frames of Mind: The Theory of Multiple Intelligences. Basic Books, New York (2011)

[9] Gartner, M.: Reflections on multiple intelligences: Myths and messages. Phi Delta Kappan **77**, 200–209 (1995). https://learnweb.harvard.edu/WIDE/courses/files/Reflections.pdf

[10] Grothendieck, A., Dieudonné, J.: Eléments de Géométrie Algébrique I. Springer (1971)

[11] Gust, H., Kühnberger, K., Krumnack, U., Schwering, A.: Analogical reasoning: A core of cognition. Zeitschrift für Künstliche Intelligenz (KI) **1/08**, 8–12 (2008)

[12] Kurzweil, R.: The Singularity Is Near: When Humans Transcend Biology. Penguin (Non-Classics) (2006)

[13] Lakoff, G., Johnson, M.: Metaphors We live By. Chicago University Press (1980)

[14] Lévi-Strauss, C.: Structural Anthropology (Revised Edition). Basic Books (2008)

[15] Mazur, B., Stein, W.: Prime Numbers and the Riemann Hypothesis. http://wstein.org/rh/rh.pdf

[16] Newell, A., Simon, H.: Gps, a program that simulates human thought. Lernende Autometen (1961). Reprint:Feigenbaum Feldman, Computers and Thought, 1963, luger

[17] Wiles, A.: Modular elliptic curves and fermat last theorem. Annals of Mathematics **142**, 443–551 (1995)