

Data Science and Intelligence

Favio Vázquez*

Abstract

I show in this article that Data Science, Big Data, and Artificial Intelligence (AI) are serious areas of research and development. We need all of them to get to Artificial General Intelligence (AGI), and also how Data Science is crucial to this endeavour.

What is Intelligence?

Not an easy question to answer. I struggled a while ago trying to define what it was, but I came across a single-phrase definition that I liked.

So let's define intelligence as: The ability to accomplish complex goals.

But what is complex here? How are we defining something complex? If you look in the Internet, you will find several different definitions; but, I think the "main" one is close to what I think the definition of intelligence is talking about.

If we think that something complex is something having many parts related to each other in ways that may be difficult to understand, then we can say that something complex is a mix of things or parts, which combine together to form a bigger thing, and the way those parts are related is not very easy to understand.

For example, something complex that has many parts which work together in a way that is not that easy to understand is a car. But, if we take a look, the individual parts are not that hard to understand. I'm not saying that they are easy to build or see exactly what they do; however, it is easier to grasp what they do.

So, we can say now that intelligence is:

The ability to accomplish difficult goals by understanding the parts that form the main goal.

These goals will be defined in the context we want, but now we want to focus on the field of Artificial Intelligence (AI). So as AI wants to build intelligence using machines and computation trying to mimic the ways we as human see, hear, learn, and more, these goals will be seeing, learning, hearing, moving, understanding, and more.

What is Understanding?

Other important concept that we need to look at is understanding. We have used that word several times by now, so let us define it:

Understanding is the ability to turn complex information into simple, useful information.

We need this, and we talked about it when we saw the parts of the car. When we are trying to understand, we decode the parts that form this complex thing, and transform the raw data we collected in the beginning to something useful and simple to see.

We do this by modeling. This is the process of understanding the "reality", the world around us, but creating a higher-level prototype that will describe the things we are seeing, hearing, and feeling; however, it's a representative thing, not the "actual" or "real" thing.

So, how we humans create intelligence: by modeling the world around us, understanding its parts, transforming the raw data we collected into useful and simple information in order to see how these parts form more complex things, accomplishing in the end goals, i.e., "difficult" goals.

Why Do We Need to Create Intelligence with AI?

I think the recipe to create intelligence is not that hard at a high level. This is why I propose we need to do it:

* Data Scientist, BBVA Data and Analytics and Chief Data Scientist at Iron AI.

Big Data + AI + Data Science = Artificial General Intelligence

I'm talking about Artificial General Intelligence (AGI) as the main goal of this revolution. AGIs are general-purpose systems with intelligence comparable to that of the human mind (or maybe beyond humans).

We need Big Data as a Catalyst to get to AGI, because with more data, in addition to new ways of analysing data and better software and hardware, we can create better models and better understanding. We need the current state of AI, very close to Deep Learning, Deep Reinforcement Learning, and its surroundings, and then we need Data Science as the controller and science behind this revolution.

What is Data Science?

This definition may cause controversy for some people, but this is something which I think is very close to what the leaders (both theoretical and in the business) are saying right now. Thus,

Data Science is the resolution of business/organization's problems through mathematics, programming, and the scientific method that involve the creation of hypotheses, experiments, and tests through the analysis of data and the generation of predictive models. It is responsible for transforming these problems into well-posed questions that can also respond to the initial hypothesis in a creative way. It must also include the effective communication of the results obtained and how the solution adds value to the business/organization.

And with this definition, we can define who a Data Scientist is. Thus,

A Data Scientist is a person (or system?) in charge of analysing business/organization's problems and

gives a structured solution starting by converting this problem into a valid and complete question, then using programming and computational tools to develop codes that prepare, clean, and analyse the data to create models and answer the initial question.

What I'm saying here is that Data Science is very much linked to the business, but it is a science in the end, or in the process of becoming one, or maybe not. I think, it could be very useful that Data Science is a Science because if that's the case, every project in Data Science should be at least:

Reproducible: Necessary for making easy to test other's work and analysis.

Fallible: Data Science and Science are not meant to look for the truth; they are meant to look for knowledge, so that every project can be substituted or improved in the future, as no solution is the ultimate solution.

Collaborative: The data scientist doesn't exist alone; he needs a team. This team will make things possible for creating intelligence and solutions. Collaboration is a big part of science, and data science should not be an exception.

Creative: Most of what data scientists do is new research, new approaches, or takes on different solutions, so their environment should be very creative and easy to work. Creativity is crucial in Science, and is the only way we can find solutions to hard and complex problems.

Compliant to Regulations: Right now, there are a lot of regulations on Science, not that much on Data Science, but there will be more in the future. Is important that the projects we are building can be aware of these different types of regulations so we can create a clean and acceptable solution to the problems.