

# Statistical Learning

Due Monday, May 15 on Moodle

## Homework-01

### General Instructions

- You can use *any* programming language you want, as long as your work is runnable/correct/readable. Two examples:
  - **In R:** it would be nice to upload a well-edited and working **R Markdown** file (`.rmd`) + its **html** output.
  - **In Python:** it would be nice to upload a well-edited and working **notebook** (or similia).

### In case of R

If you go for R, to be sure that everything is working, start **RStudio** and create an empty project called **HW1**. Now open a new **R Markdown** file (**File > New File > R Markdown...**); set the output to **HTML mode**, press **OK** and then click on **Knit HTML**. This should produce a **html**. You can now start editing this file to produce your homework submission.

- For more info on **R Markdown**, check the support webpage: [R Markdown from RStudio](#).
- For more info on how to write math formulas in LaTeX: [Wikibooks](#).

---

## Part II: Era of the Bots

No wonder that **ChatGPT** and all its LLM-siblings are big these days, and probably even your beloved grandpa is asking you some amazingly emotional questions about it. I've already shared some opinions on this, [did I?](#)

Now, although we can look at things like the recently announced **Dolly 2.0**, a **ChatGPT** competitor for the people, as a cool way to “**democratize**” AI tech, a question naturally rises up from the ashes of the (irrational) fear that, undoubtedly, these tools generates in large slices of our society: do we (again, as a society) really need this type of over-bloated “**democratization**” or it's about time to lower the hype and focus on **regulation**?

**Just as an example**, the U.S government agency responsible for these things is currently **seeking public comments** to help it put together laws to protect us from the distopic, **Black Mirror**-like horror future that literally every “evil programmer” (a.k.a. *you*) is cooking up on her/his laptop right now.

### ~ Q1 ~

Within your team of *data-people* that, in all likelihood, will also look for a *data-heavy* job in the near future, is there a informed opinion, or simply an emerging sentiment, on these still controversial aspects of the field you are swimming into? Do you feel empowered by the rising of *these*, possibly **aligned**, bots? And, by the way, *aligned*, okay, but to *what* or *who*? Please, share your honest (a.k.a. not bot-generated) thoughts and (irrational?) feelings, also citing/quoting all the relevant material/sources you used/read/agree-disagree with (included **mine**). Don't be lazy, don't be shy, go for it!

All right, **ChatGPT** got all the rage but, of course, there's even more out there, right?

Introducing **AutoGPT**, a “tiny” **recent variation on the LLM idea** that turn our preferred chatbot into a **semi-autonomous agent**: when you give it a goal, instead of giving you a response and then waiting for prompt number two, it will give itself *feedback* based on its *own* performance and continue acting in pursuit of its goal until you shut it down **Terminator**-style.

**AutoGPT** spawned many similar projects many of which were trending on **GitHub** a few days ago, including an official Microsoft project called **Jarvis**. In addition, **Park et al.** published a study where 25 of them were put in a video game and when one of them was given the idea to throw a Valentine's day party, the agents autonomously sent out invites, met new acquaintances, asked each other on dates to the party (... and then actually showed up :).



## Q2

1. For this question you *should* build **federations** of up to 3 teams (the team-leaders list *is available on Moodle*), that I will consider as small **startup companies**.

Hence, the first things you will provide to answer this question are: **(A)** the list of teams involved in the company; **(B)** the company name, motto and possibly its logo.

2. You should think about this exercise as the design phase in the development of a new software product that I (i.e. the *customer*), am commissioning to you.

The second phase, in which you will develop a *veryveryvery* alpha version of your product, will be part of **Homework 2**. Hence, the basic requirement your proposal has to fulfill is: “doability” in under a month.

3. Finally, the actual task: chatbots and alike are sparking much attention and interest in different fields like **health**, **medicine** and **psychology**, just to mention a few.

As you might know, I am all-in with Education + Statistics, and to **hire a comedian** is really not an option...

Hence, I would like your company to come forward with a (surely innovative!) **GPT-based** companion to teach and/or practice some relevant concepts and/or tools in Prob/Stat/Stat-Learn (you choose) to high-school students or 1<sup>st</sup> year college students → do not necessarily think to a MATH/STAT/CS/etc student, also students from **humanistic disciplines** may be your target.

Of course, feel free to plan the inclusion of **OpenAI API** (**R**, **Python**), or any other *chat-auto-boty-stuff* of your choosing, at least as long as no real money will ever **be part of your proposal**.