Software & digital practices at CREST

Émilien Schultz - Alexis Guyot - Claire Ecotière - Philippe Pinczon du Sel

2024-10-03

The survey was conducted between July 2024 and september 2024 among CREST researchers on their digital practices. This summary higlights the main results.

A total of **89** respondants answered the survey (**81** completed it totally, with a mean duration of **14** minutes ; only some question were mandatory so the total number of respondants can vary).

Respondents profile

Respondants are mainly ENSAE researchers (**92%**), mostly faculty/phd student, with a high proportion of economists (**61%**)

		N	%
Organization	ENSAE		
	ENSAI		
	Total		
Field	Computer scien		
	Economics		
	Finance		
	Other		
	Sociology		
	Statistics		
	Total		
Status	Faculty		
	Other		
	Phd student		
	Postdoctoral re		
	Research assist		
	Total		
Seniority	Between 1 and		

	N	%
Between 5 and		
Less than 1 yea		
More than 10 y		
Total		

Digital practices

Almost all respondants (**87%**) are involved in digital processing of data (including simulations) : **99%** are using numeric data, **51%** textual data, **14%** images and **4%** audio. More specifically, **29%** uses experimental data.

There is a wide diversity of practices at CREST, both regarding computing activites or storage. More than half of the respondant uses dataset around 1 Gb or less (**52%**), and only **7%** declared to use a dataset bigger than 100 Gb (**10%** reported to not know).

Overall, **83%** reported to have enough computing ressources and **77%** they had enought storage. Nevertheless, a few respondents reported the limit of available ressources.

> My research is in computational statistics and machine learning. You can't seriously compete with other teams in this field without access to big clusters (100s of CPU cores, GPUs).

The question of getting enough GPU memory (VRAM) for LLM was mentionned a few time. Comments were made on the necessity of flexible cloud storage.

The diversity of practices is visible on the hardware used. For instance, the distribution for the question *'on which computers do you perform these data processing tasks/computations ?'* :

	[Locally, with my office computer]		[Locally, on my laptop (GENES/CREST or personal)]		[Locally, on a c
	N	%	N	%	N %
No	48.0	57.8	26.0	31.3	79.0
Yes	35.0	42.2	57.0	68.7	4.0
Total	83.0	100.0	83.0	100.0	83.0

And for the question *'Where do you currently store your data ?'*

	[Locally, with my office computer]		[Locally, on my laptop (GENES/CREST or personal)]		[Locally, on a c
	N	%	N	%	N %
No	55.0	66.3	30.0	36.1	80.0

	[Locally, with my office computer]		[Locally, on my laptop (GENES/CREST or personal)]		[Locally, on a c
	N	%	N	%	N %
Yes	28.0	33.7	53.0	63.9	3.0
Total	83.0	100.0	83.0	100.0	83.0

Software practices

Regarding operating systems, the majority (**70%**) uses Windows, on third MacOs (36%) and only 13.5 uses Linux.

A majority of respondants (58%) reported to need a desktop computer.

The main softwares deemed to be necessary are Stata, R, Python, Matlab, Dropbox and Latex.

Only 15% of the respondants are paying software with their research funds. Some examples are : databases access, chatGPT, MaxQDA… More (43%) are paying software with their own pocket money : Dropbox, Claude/ChatGPT, Zotero storage, Overleaf, Dropbox. To note : some of those software are available in the laboratory offer.

To the question on the needs, several suggestions were made : chatGPT, Acrobat Pro, Dropbox/Google drive cloud storage, OCR software, or Premium Overleaf account

The current programming practices at CREST shows the diversity of languages, with a dominant of R, Python and Stata.

			r		python		julia		stata		matlab		sas			
			N	%	N	%	N	%	N	%	N	%	N	%		
No	30.0	36.1	25.0	30.1	72.0	86.7	39.0	47.0	52.0	62.7	73.0	88.0				
Yes	53.0	63.9	58.0	69.9	11.0	13.3	44.0	53.0	31.0	37.3	10.0	12.0				
Total	83.0	100.0	83.0	100.0	83.0	100.0	83.0	100.0	83.0	100.0	83.0	100.0				

A lot of software paid from researcher's poket

What are the software needed ? Few demands

TO DEVELOP

A vast majority of respondants (**74%**) are using generative model for their research, one third reported to use it a lot. Three out of for are using free access solutions.

This use cover a diversity of tasks :

- editing english