Scrolling Through *Slop*
Quantifying AI video generation on Sora

Tristan Coignion[1], A. Bugeau[1,2], G. Guennebaud[1,3], AL. Ligozat[4]
1. Univ. Bordeaux, CNRS, Bordeaux INP, LaBRI, UMR 5800, Talence, France - 2. IUF - 3. Inria
4. Université Paris-Saclay, CNRS, ENSIIE, LISN, 91400, Orsay, France

Find more online!
tristan.coignion.fr/posts

Our goal:
Anticipate the potential environmental impacts of AI video generation.

→ How to model the impacts of AI videos' life cycle? → Gather insights with open-source video generation models.

→ How to build prospective and baseline scenarios? → What is the current usage (TO) of AI video generation?

This poster →

How we did it

Sora's feed is chronologically ordered.
↳ **Easy to scrape**

6 bots scraping the feed
+
10 bots scraping video metadata
↓
Scraped 1.33M videos over 18 days.

~74,000 public videos generated daily on sora.chatgpt.com

← That's 5.81 TiB every month!

Misc. stats
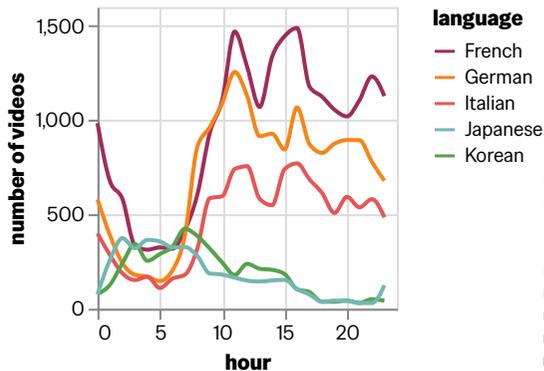

users / videos per user
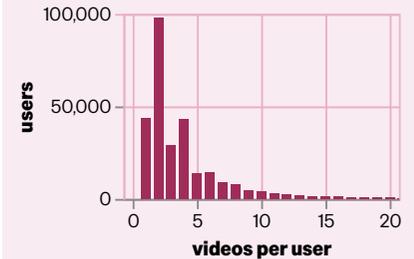
75% of videos last 5s; 20% last 10s

80% of the videos are in 480p

Only 0.8% of videos get one or more likes after 7 days

When is it used?
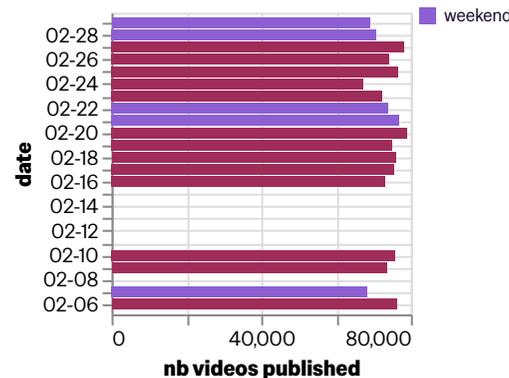
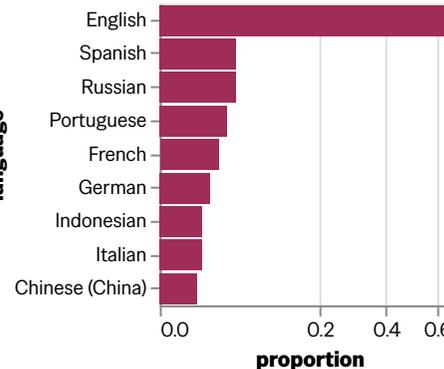Generations happen mostly during working hours, and greatly decrease at night.

number of videos / hour

language: French, German, Italian, Japanese, Korean

Does usage vary over time?

Rate of generations are relatively stable across days and weekends.

date / nb videos published — weekend

Who uses it?

While prompts are mostly written in English, Sora seems to be used all over the world.

language: English, Spanish, Russian, Portuguese, French, German, Indonesian, Italian, Chinese (China) / proportion

---

Open questions

What about generations from the API and competitors?

What are the videos used for?

How will usage evolve over time?