#CONF'

"Le RAG : booster de connaissances pour la GenIA ? "

Barbara DELACROIX
& Marvin SANT

Fondateurs @ Devana.ai

devana.ai

Le 19/12/23 à 19h

le wagon


First time, huh?

Schedule

 1- News (15min)

 2- Talk (1h30)

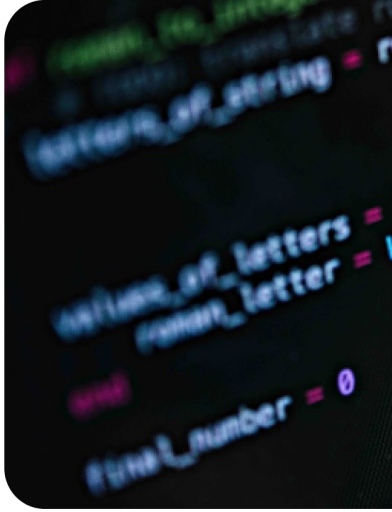
 3- Enjoy ;)

Sponsor turbo-platinum

❤️

Screeb

Accélérez votre carrière.
Formez-vous aux métiers de la tech.



Notre vision

Nous croyons que chacun(e)
peut s'épanouir continuellement
dans sa carrière.

Reconversion

Bootcamps intensifs en Web & Data
400h

Montée en compétences

Skill Courses à temps partiel
40h

À propos du Wagon

Le Wagon est un des leaders mondiaux des formations immersives tech depuis 2013.

- **Débutez une nouvelle carrière**
- **Obtenez des compétences techniques**
- **Cours en ligne ou sur campus**
- **Formats temps plein et temps partiel**
- **Formations Data, Développement Web & No-code**

23,000+
Diplômés dans le monde

#1
La formation la plus reconnue

4,98/5
Sur +6,600 avis

40+
Villes sur tous les continents

200+
Startups fondées par nos alumni

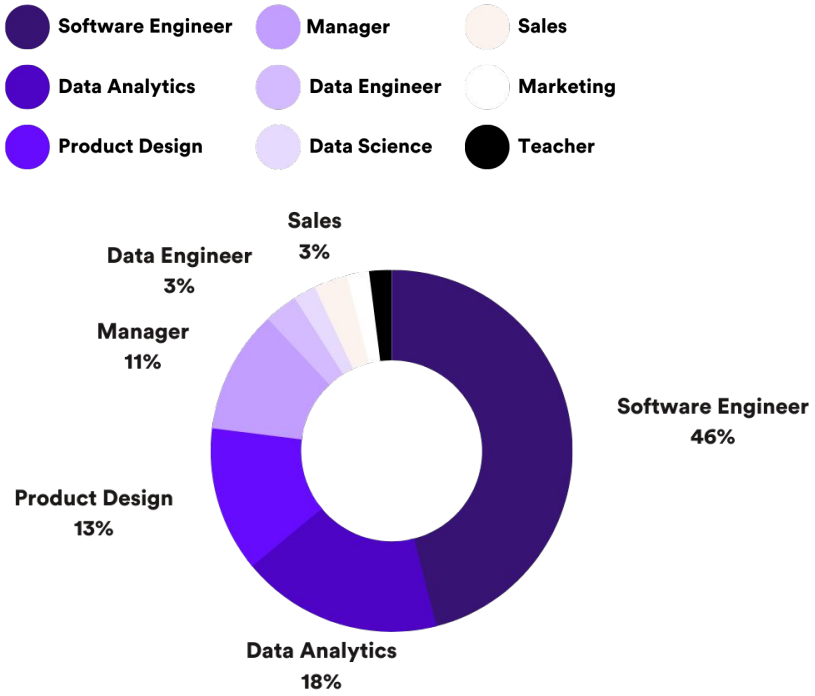
$1B
Levé par nos alumni

Présence mondiale

Campus du Wagon

L’employabilité après nos formations

Les données sont plus parlantes que les mots.



34 jours

En moyenne pour décrocher un emploi dans le secteur de la Tech en Europe.

93%

De nos diplômés des campus Européens ont trouvé un job dans une entreprise Tech, se sont lancés en Freelance ou ont lancé leur propre startup.

45k€

Salaire médian en Europe pour leur premier emploi post-bootcamp.

le wagon

Nos bootcamps 200-400h

Pour accéder à de nouvelles opportunités grâce aux formations immersives en développement web et en data.

- ✓ **1 professeur pour 7 étudiants**
- ✓ **Temps partiel ou Temps plein (de 2 à 6 mois)**
- ✓ **Sur campus ou en ligne**

Développement Web

Devenez développeur.se et créez des applications web, de la base de données à l'interface utilisateur.

Data Science & IA

Obtenez les compétences fondamentales du Data Scientist et créez vos propres modèles d'IA.

Data Analytics

Apprenez à transformer vos données en informations exploitables et décrochez un poste de Data Analyst.

Data Engineering

Concevez des pipelines de données et développez des bases solides pour lancer votre carrière en tant que Data Engineer.

le wagon

Notre methodologie

- 1 professeur(e) pour 7 étudiants
- Une classe à taille humaine
- Des professeurs passionnés pour vous aider
- Travaillez sur de vrais projets de startups
- Ajoutez votre projet final à votre portfolio
- Une pédagogie basée sur la pratique

★★★★★

"Si vous cherchez une formation qui offre un programme complet, un environnement d'apprentissage formidable et un incroyable soutien professionnel post-bootcamp, alors Le Wagon est fait pour vous !"

Joseph Gulay
Désormais Data Analyst
Ernst & Youngt

4.98 / 5
2301 reviews

4.98 / 5
2259 reviews

4.9 / 5
714 reviews

Choisissez le rythme qui vous convient le mieux

Sur campus ou en ligne, à temps plein ou à temps partiel !

Temps plein

2 mois

Prêt(e) à plonger dans la Tech ? Rejoignez notre programme immersif. Du lundi au vendredi, de 9h à 18h.

- **Un diplôme en 2 mois**
- **Apprentissage en groupe toute la journée**
- **40 heures d'apprentissage par semaine**
- **Sur campus ou en ligne**

Temps partiel

6 mois

Vous avez des contraintes d'emploi du temps ? Apprenez pendant votre temps libre grâce au temps partiel flexible.

- **Un diplôme en 6 mois**
- **Apprentissage en groupe, en ligne**
- **16 heures d'apprentissage par semaine**
- **En ligne ou sur le campus de Paris**

Alternance

16 mois

Vous aimeriez vous former tout en étant rémunéré(e) ? Rejoignez nos formations 100% financées en alternance.

- **Un diplôme en 16 mois**
- **Apprentissage ou Professionnalisation**
- **3 mois de formation sur campus**
- **12 mois en entreprise pour pratiquer**

le wagon

Le Wagon s’associe à Google pour former les futurs talents de la Data & l’IA

En rejoignant l’un de nos bootcamps Data, nos étudiants ont la possibilité d’intégrer le programme “Bootcamp Numérique IA” développé avec Google.

- Un accompagnement d’excellence
- Le consortium d’employeurs de Google
- Des certifications reconnues*
- Votre formation 100% financée*

En savoir plus

*offre réservée aux étudiants éligibles

Nos skill courses 40h

(nos formats courts)

Pour gagner en performance dans son job actuel ou découvrir un nouveau sujet.

✔ **Temps partiel** ✔ **Session live en ligne**

Data Analytics

Web Analytics & Tracking →

Data Analytics Essentials →

No-code / Low-code

Growth & Data Automation →

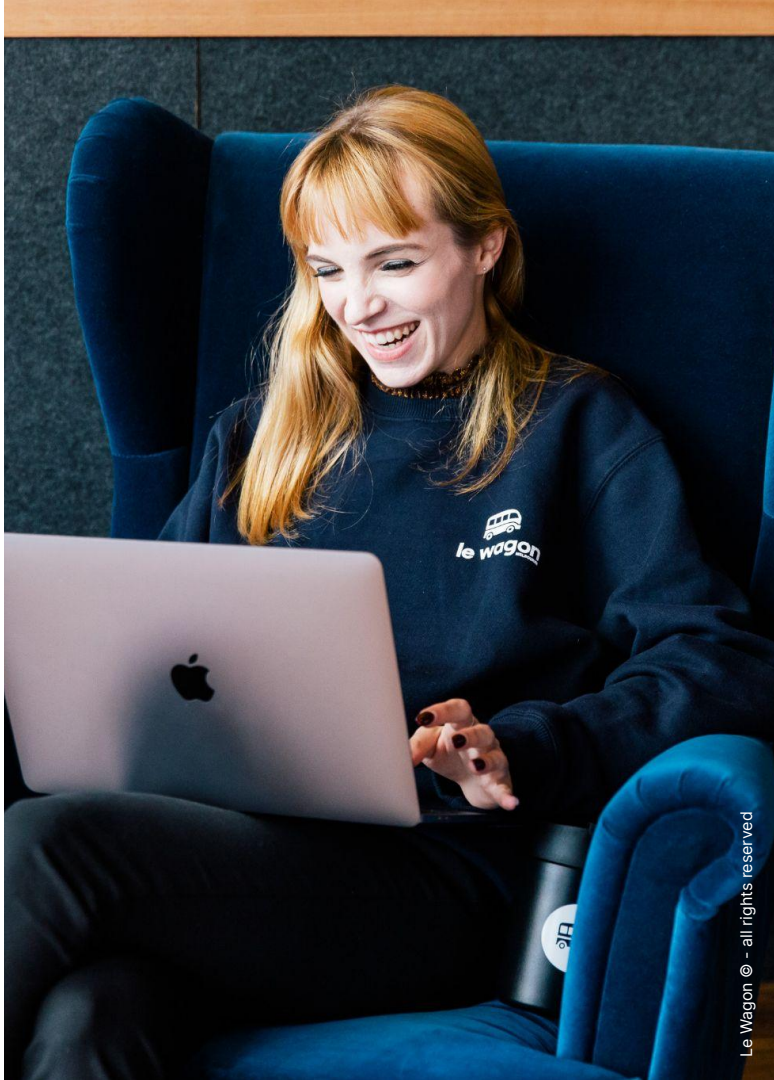
Web Design & Webflow →

Product Dev. & Airtable →

Data Science

Python & Machine Learning →

le wagon



Tout le monde peut bénéficier de financement.

Maintenant que vous envisagez de faire vos premiers pas dans le monde de la Tech, vous vous demandez peut-être comment financer votre formation ?

Options de financement

Alternance

Formation **100% prise en charge** et **rémunération** pendant toute la durée du contrat (16 mois)

Financements personnels

Paiement en **3 à 12 fois** avec Alma, prêt BNP Paribas

Financements publics

Transition Pro, AIF, CPF, Aides régionales, POEI... **De nombreuses aides existent selon votre profil**.

Tarif des formations

- **Skill Courses : de 1490€ à 1690€**
- **Bootcamps : de 5900€ à 8900€**

Nos prochaines sessions

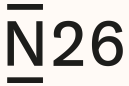
Sur le campus de Nantes

- **Développement Web (Temps plein)**
  - → Du 15 janvier au 15 mars 2024 ❄️
  - → Du 15 avril au 14 juin 2024 🌸
- **Data Analytics (Temps plein)**
  - → Du 15 janvier au 15 mars 2024 ❄️
  - → Du 15 avril au 14 juin 2024 🌸

En ligne

- **Bootcamps à temps plein**
  - → Du 29 janvier au 29 mars 2024 ❄️
  - *Développement Web, Data Analytics, Data Science*
- **Bootcamps à temps partiel Flexible**
  - → Du 10 février au 27 juillet 2024 ❄️
  - *Développement Web, Data Science*
- **Formations courtes Skill Courses**
  - → Du 23 octobre au 14 décembre 2023 🍁
  - → Du 15 janvier au 7 mars 2024 ❄️

Ces entreprises recrutent nos alumni



Envie d’en savoir plus ?

- **Votre situation / projet,**
- **Contenu des formations,**
- **Processus d’admission,**
- **Insertion professionnelle,**
- **Solutions de financement,**
- **Etc.**
- Posez-moi toutes vos questions sur les formations du Wagon lors d'un RDV informel !

Valentin Napoli

Directeur de campus

Le Wagon Nantes & Rennes

06 79 99 24 96

valentin.napoli@lewagon.org

/valentinnapoli

le wagon

Le Wagon vous offre un cours gratuit d'introduction à l'IA !

Inscrivez-vous ce soir pour bénéficier de cette opportunité !
🚀

Introduction à l'intelligence artificielle

- ✔ 8 heures de cours
- ✔ Outils gratuits uniquement

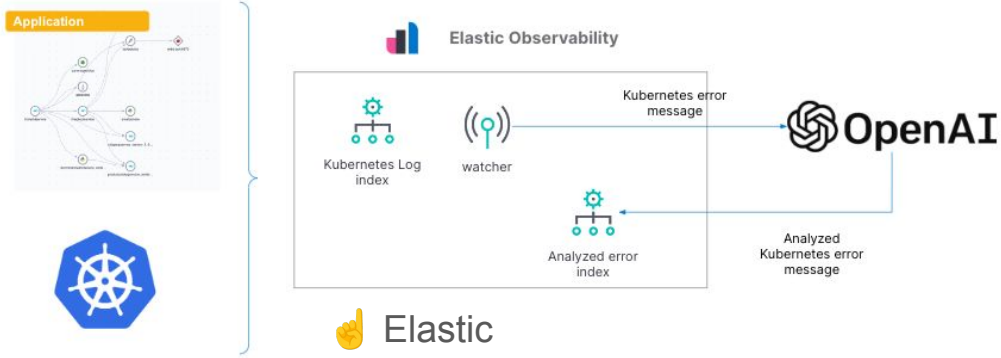
Sponsors 2024

1y partnership

For 🍕 and 🍻

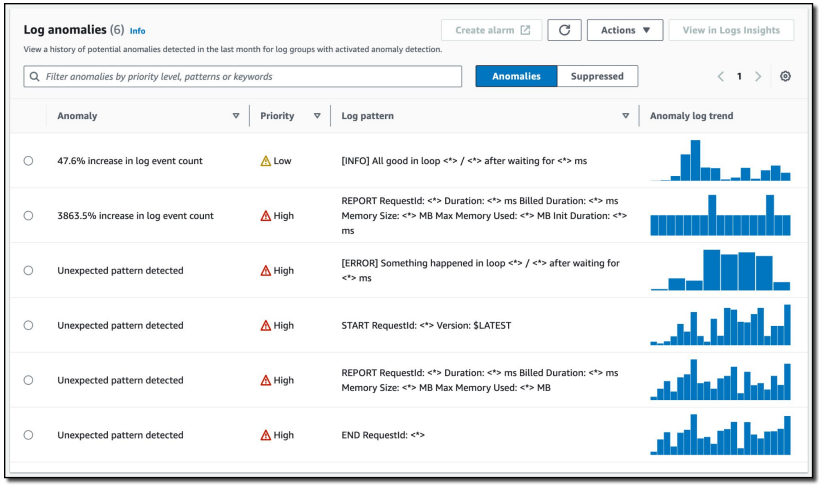
Contact us ;)

Observability: anomaly detection



☝ Elastic

👇 AWS



Resources and Maybe Solutions

Hide Details

AI SOLUTION EXPERIMENTAL

Hide Suggestion

👇 Sentry

Hey there! 👋

Problem Description

Looks like your app is hanging for at least 2000 ms! 😱 The issue happened at `-[NSBigMutableString getCharacters:range:]` in your app code. The app is stuck and unresponsive to user input, which is not good news! 🚫

Proposed Solution

To fix this issue, you need to identify the root cause of the app hanging. You can start by profiling your app with a tool like Instruments to see if there are any performance bottlenecks. You can also review the code at `-[NSBigMutableString getCharacters:range:]` to see if there are any optimizations that can be made.

Once you've identified the root cause of the issue, you can make the necessary code changes to resolve it. The solution should be implemented in the affected method or class.

To prevent future occurrences of this issue, it's important to follow best practices for app performance optimization. This includes minimizing expensive operations, using background threads for long-running tasks, and avoiding blocking the main thread.

What Else

Hang in there! 💪 App performance issues like this can be frustrating, but with a little bit of investigation and optimization, you can get your app running smoothly again!

Here's a hip hop rhyme to help you remember this error:

```
App hanging like a chandelier,
2000 milliseconds of fear,
Check your code and optimize,
To make your app shine and rise!
```

K8S 👉

K8SGPT KUBERNETES SUPERPOWERS

code size 273 kB | build passing | release v0.3.23 | openssf best practices passing | Documentation

`k8sgpt` is a tool for scanning your Kubernetes clusters, diagnosing, and triaging issues in simple English.

It has SRE experience codified into its analyzers and helps to pull out the most relevant information to enrich it with AI.

Out of the box integration with OpenAI, Azure, Cohere, Amazon Bedrock and local models.

Mixtral 8x7B & Mixtral 8x7B Instruct

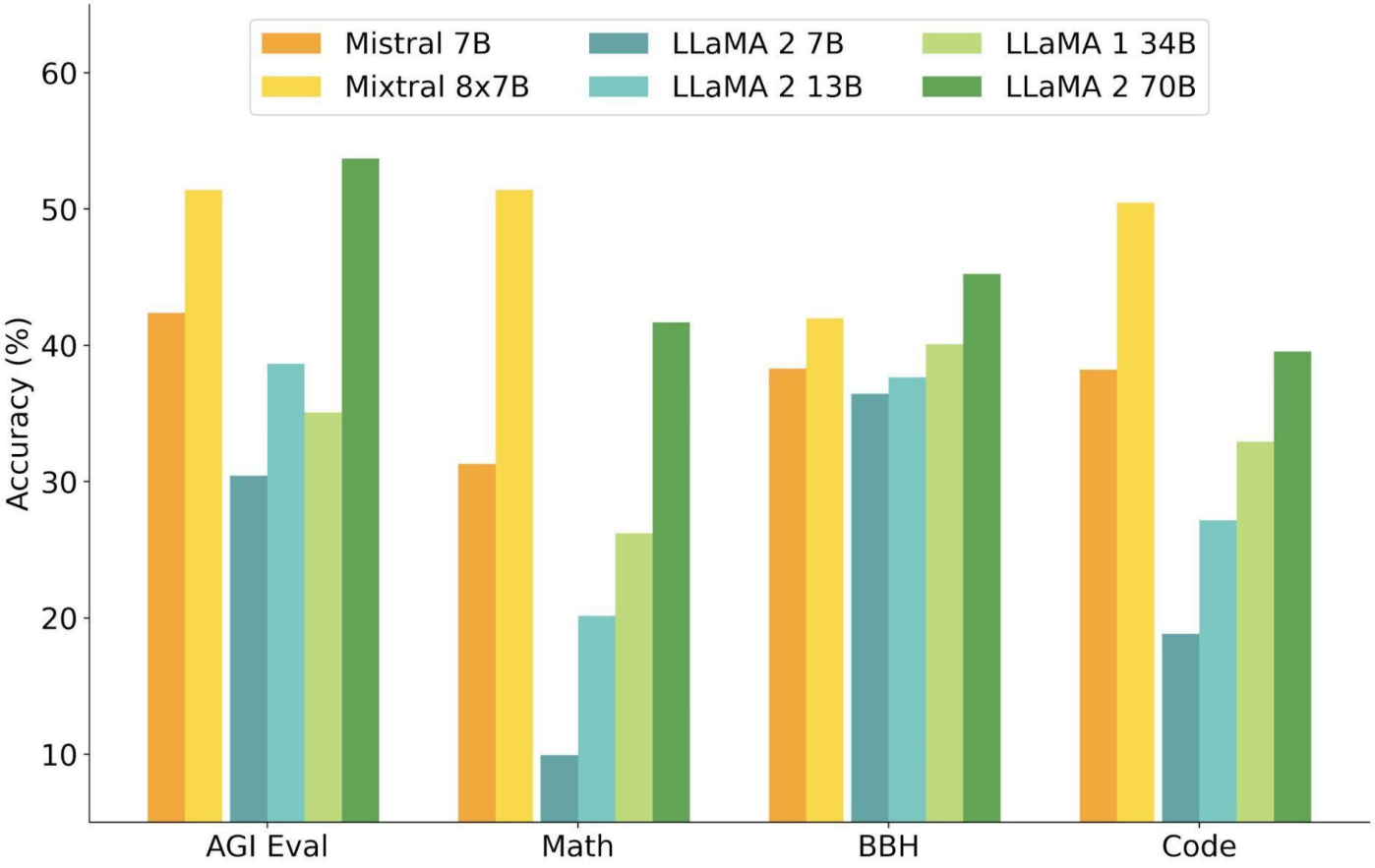
Mixture-of-Experts (MoE)

Outperforms Llama 2 70B and GPT3.5

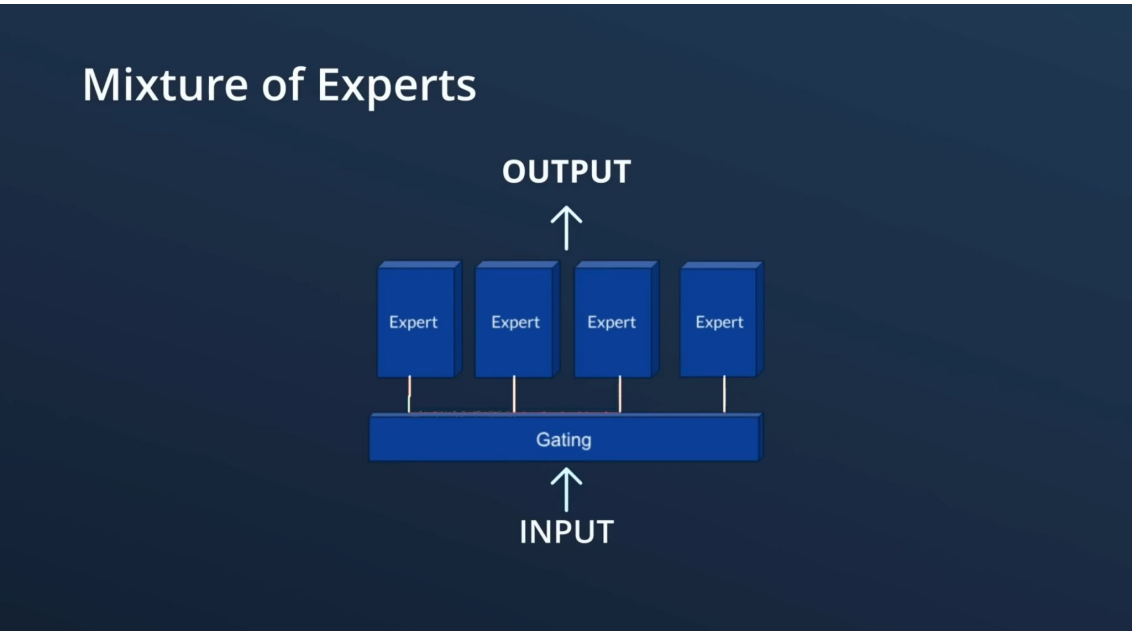
32k tokens

English, French, Italian, German and Spanish

	LLaMA 2 70B	GPT - 3.5	Mixtral 8x7B
MMLU (MCQ in 57 subjects)	69.9%	70.0%	**70.6%**
HellaSwag (10-shot)	87.1%	85.5%	86.7%
ARC Challenge (25-shot)	85.1%	85.2%	**85.8%**
WinoGrande (5-shot)	**83.2%**	81.6%	81.2%
MBPP (pass@1)	49.8%	52.2%	**60.7%**
GSM-8K (5-shot)	53.6%	57.1%	**58.4%**
MT Bench (for Instruct Models)	6.86	**8.32**	8.30


Mistral 7B
Mixtral 8x7B
LLaMA 2 7B
LLaMA 2 13B
LLaMA 1 34B
LLaMA 2 70B
Accuracy (%)
60
50
40
30
20
10
AGI Eval
Math
BBH
Code

Mixture-of-Experts (MoE)

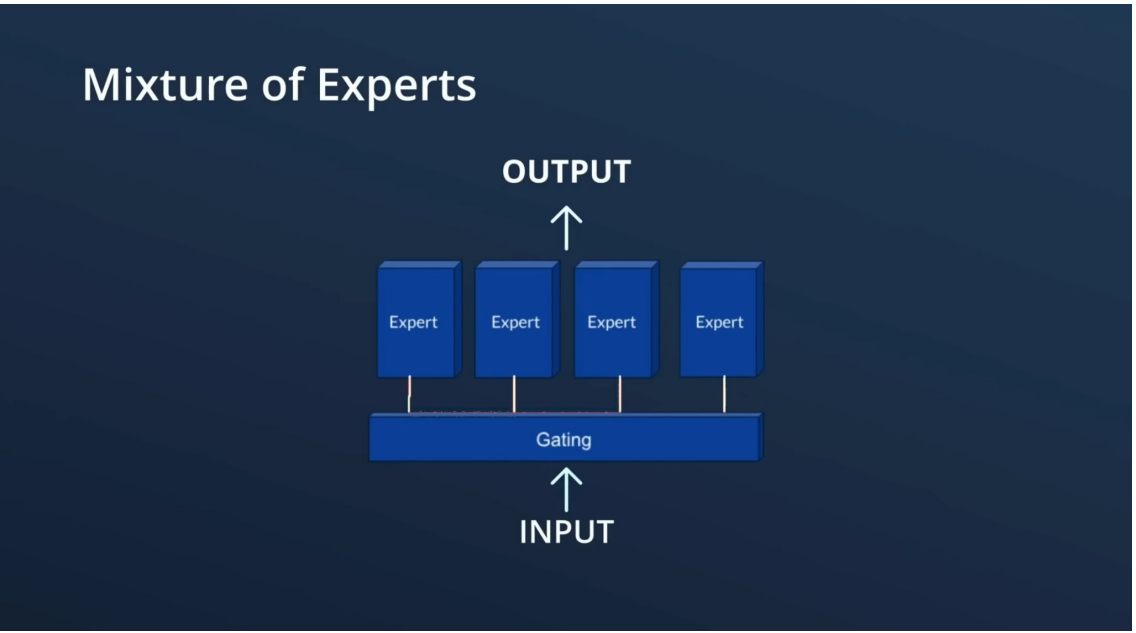


7B x 8 == 47B 😁

46.7B total parameters but only uses 12.9B parameters per token (2 experts).

Same speed and cost as a 12.9B model.

Mixture-of-Experts (MoE)



More complex fine-tuning

Unequal load balancing

Let’s play a game: inference cost reduction

JJ — oss/acc
@JosephJacks_

Last week @MistralAI launched pricing for the Mixtral MoE: $2.00~ / 1M tokens.

Hours later @togethercompute took the weights and dropped pricing by 70% to $0.60 / 1M.

Days later @abacusai cut 50% deeper to $0.30 / 1M.

Yesterday @DeepInfra went to $0.27 / 1M.

Who’s next ??? 📉

Let’s play a game: inference cost reduction

JJ — oss/acc
@JosephJacks_

Last week @MistralAI launched pricing for the Mixtral MoE: $2.00~ / 1M tokens.

Hours later @togethercompute took the weights and dropped pricing by 70% to $0.60 / 1M.

Days later @abacusai cut 50% deeper to $0.30 / 1M.

Yesterday @DeepInfra went to $0.27 / 1M.

Who’s next ??? 📉

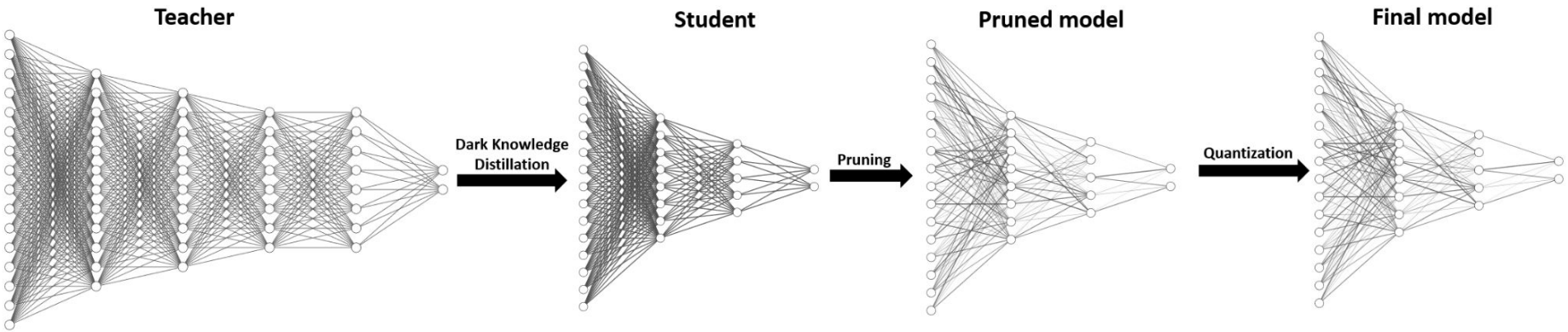
vs GPT-3.5-turbo cost: $2/1M

Mistral-medium on the road to GPT-4

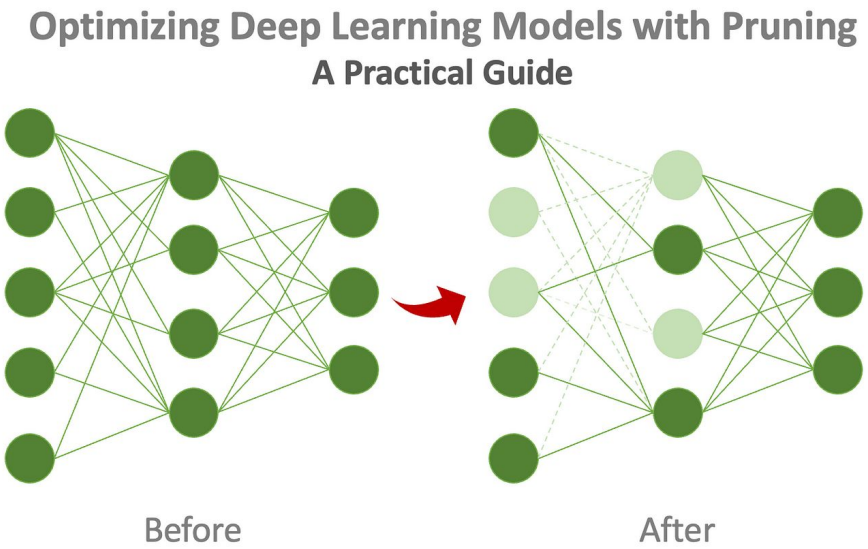
Chat Completions API

Model	Input	Output
`mistral-tiny`	0.14€ / 1M tokens	0.42€ / 1M tokens
`mistral-small`	0.6€ / 1M tokens	1.8€ / 1M tokens
`mistral-medium`	2.5€ / 1M tokens	7.5€ / 1M tokens

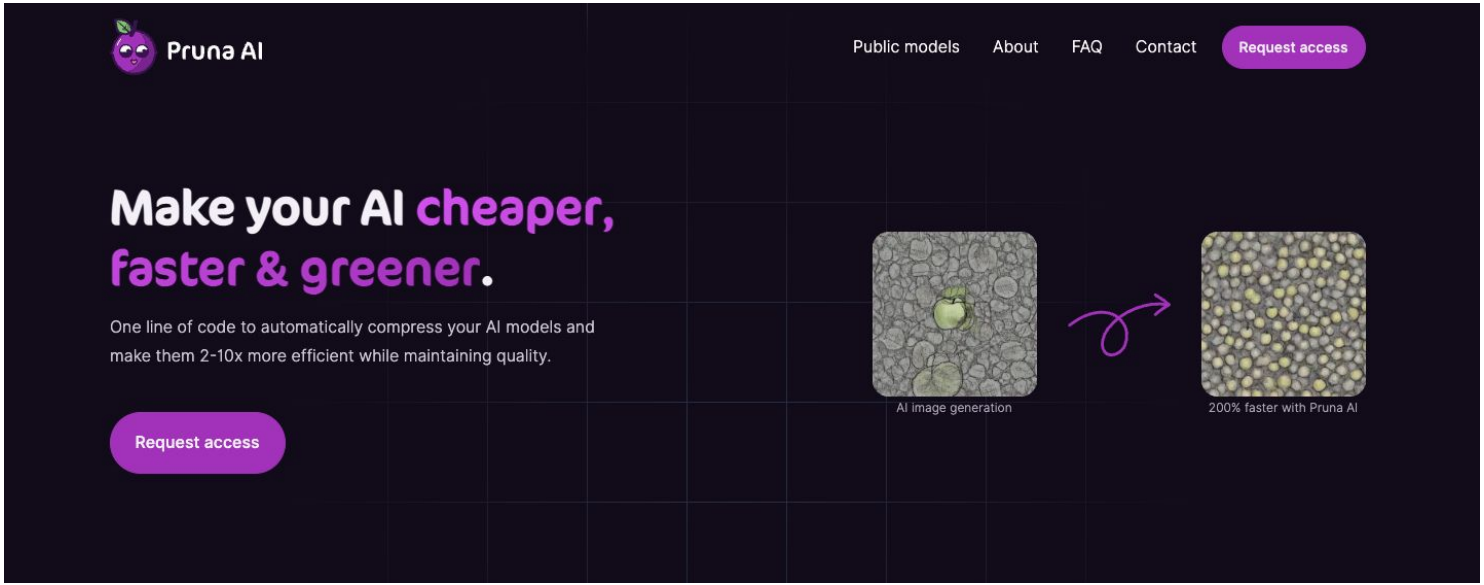
Inference & compression



Model pruning



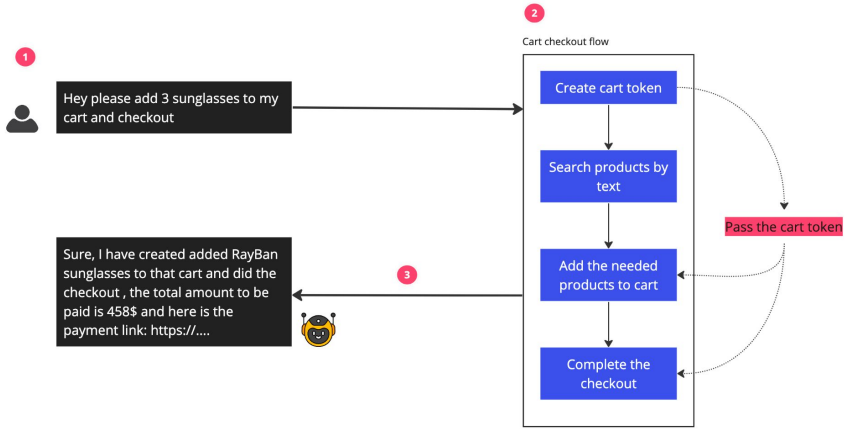
Model pruning



15% to 75% weight reduction - x2 faster

OpenCopilot

https://docs.opencopilot.so/introduction



Gladia: whisper-zero

Back

Jean-Louis Queguiner ✓
@JiliJeanlouis

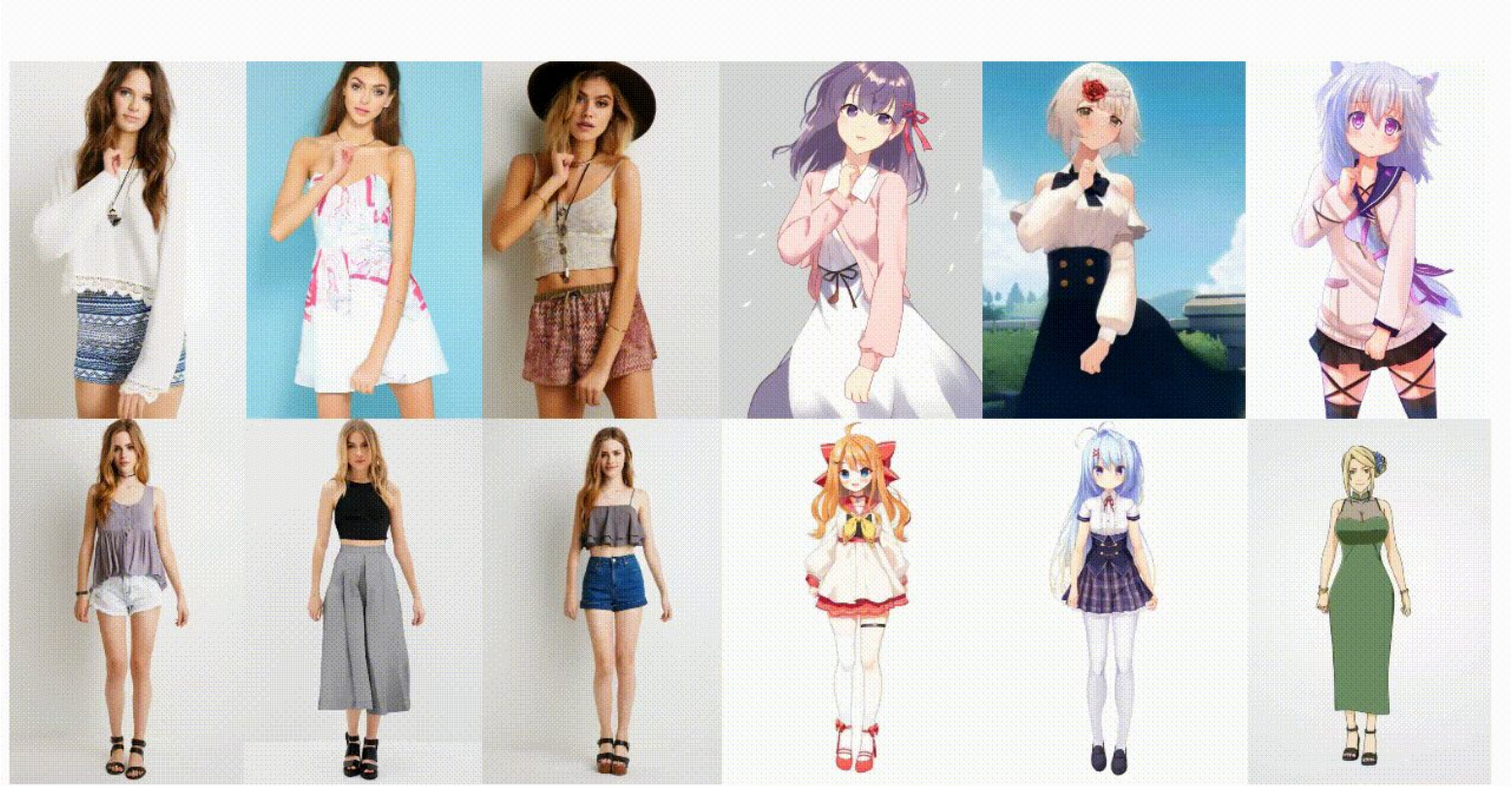
Exciting news!

After months of hard work at @gladia_io, we're thrilled to introduce Whisper-Zero, our latest model that solves major pain points.

We asked our customers what their biggest problems were:
- **hallucinations** (which we fixed ✅ 0 hallucinations)
- **language detection** (which we highly improved ✅ in the 0's % error)
- **code-switching** (change the language on the fly ✅ 0 setup needed)
- **complex environment sensibility** ✅
- **huge errors in numbers** leading to millions of $ lost in our customers' critical business process automation ✅

It leverages the Whisper architecture, extensively optimized on over **1.5 million hours of diverse audio recordings captured under various conditions, including hostile environments.**

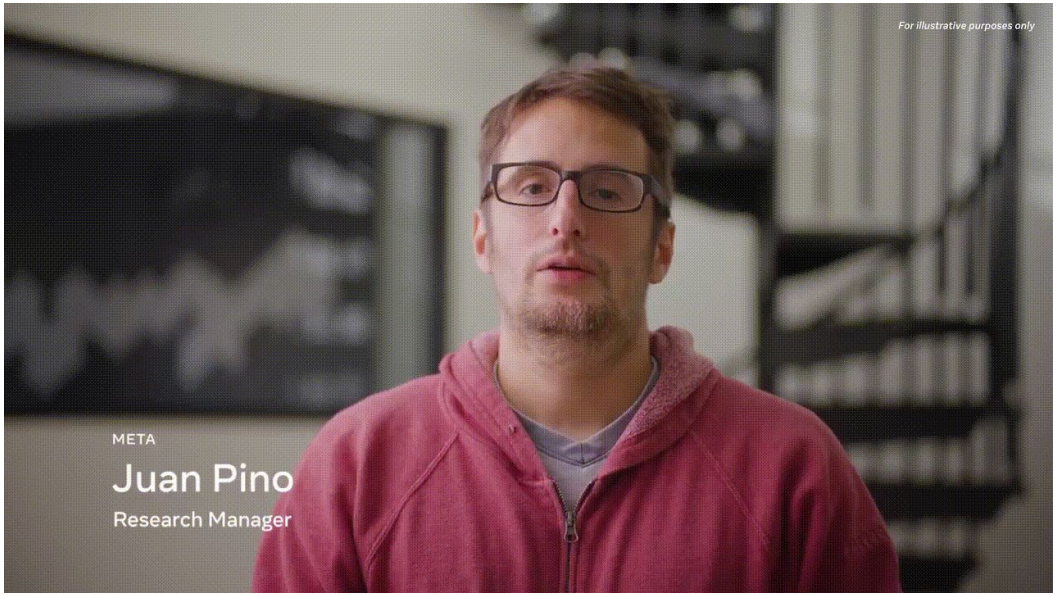
About fake demo (Alibaba)



About fake demo (Alibaba)

About fake demo (Google Gemini)

Real-time translation



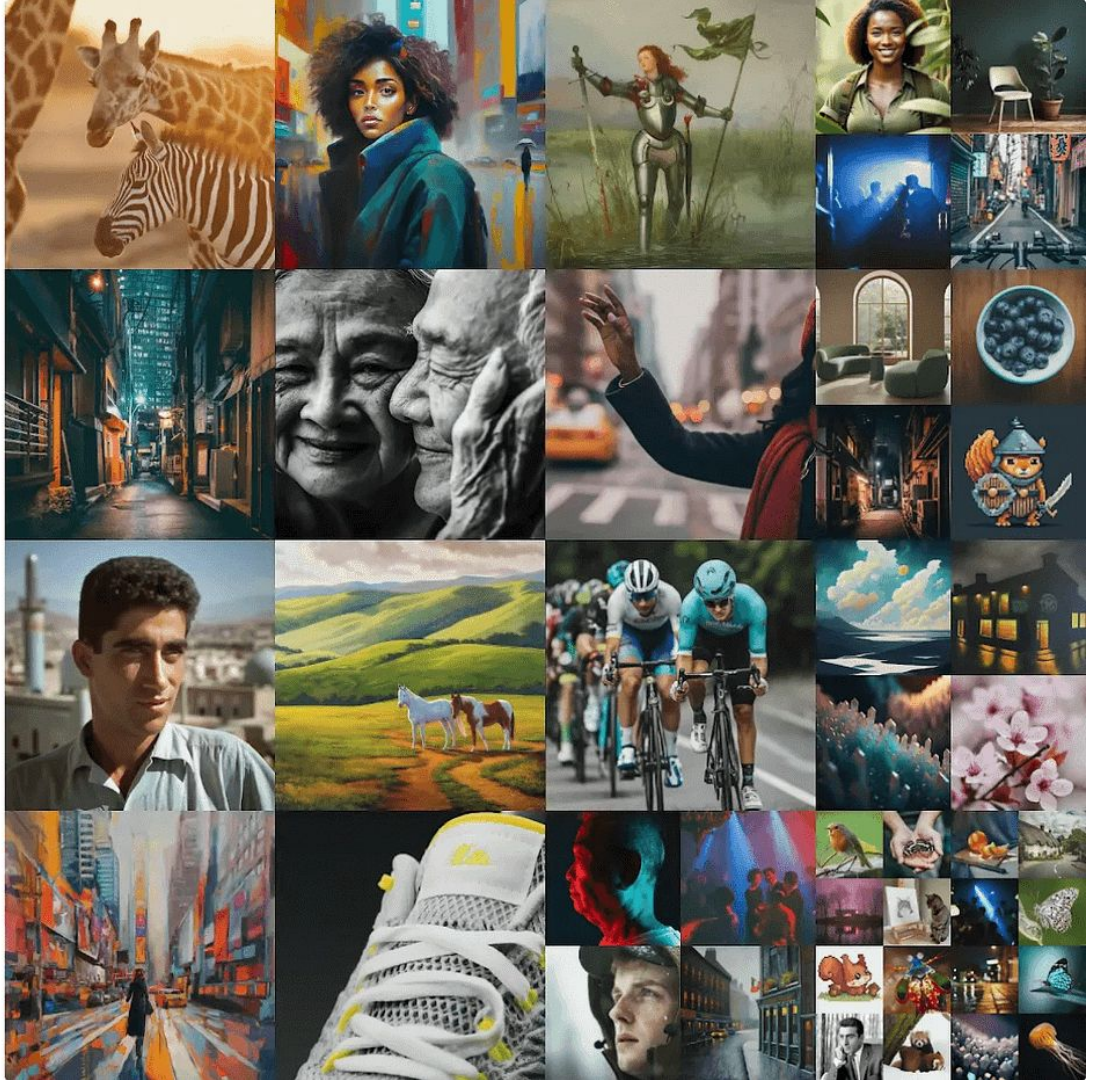
Real-time (2s latency)

Output before end of sentence.

Open-source weights

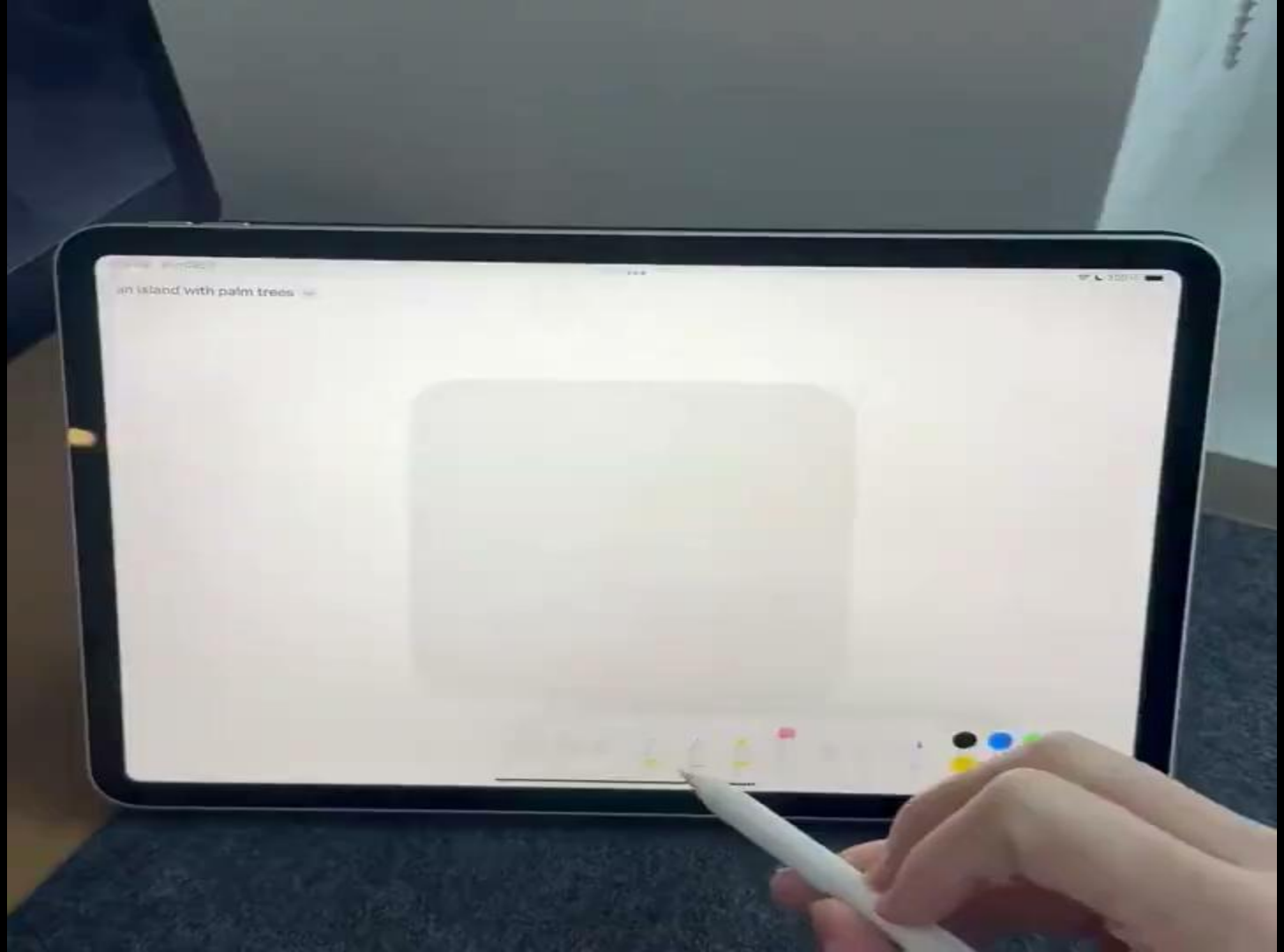
https://x.com/BrianRoemmele/status/1735836569960173602?s=20

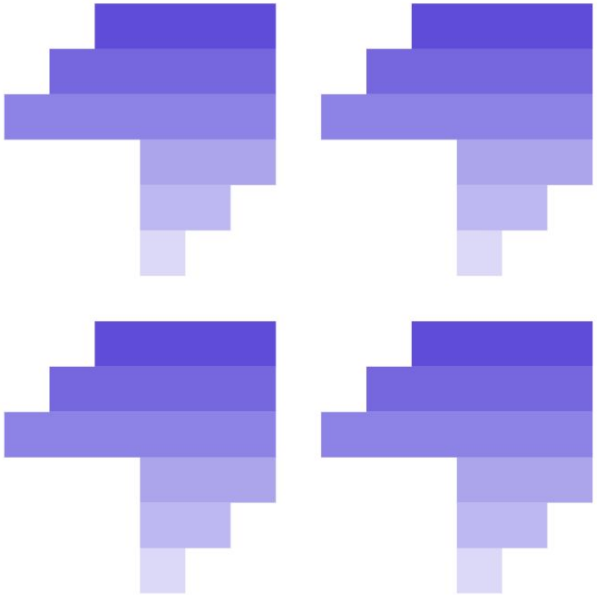
Imagen 2 most advanced text-to-image technology



Google Cloud

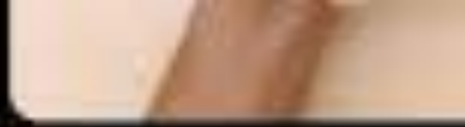
MAGNIFIC AI


an island with palm trees

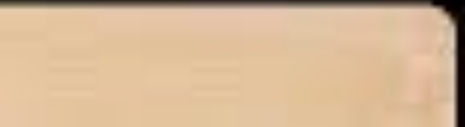
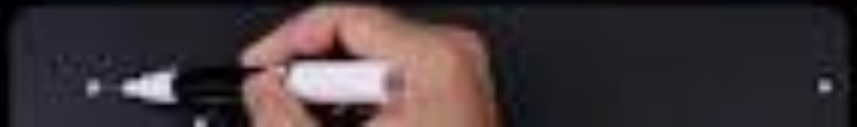


fal







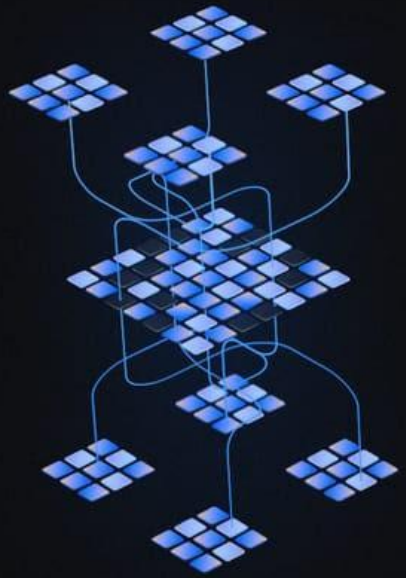




Sagittarius

localhost:5173

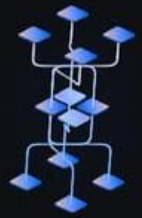
Stop

Gemini



Ultra

Pro

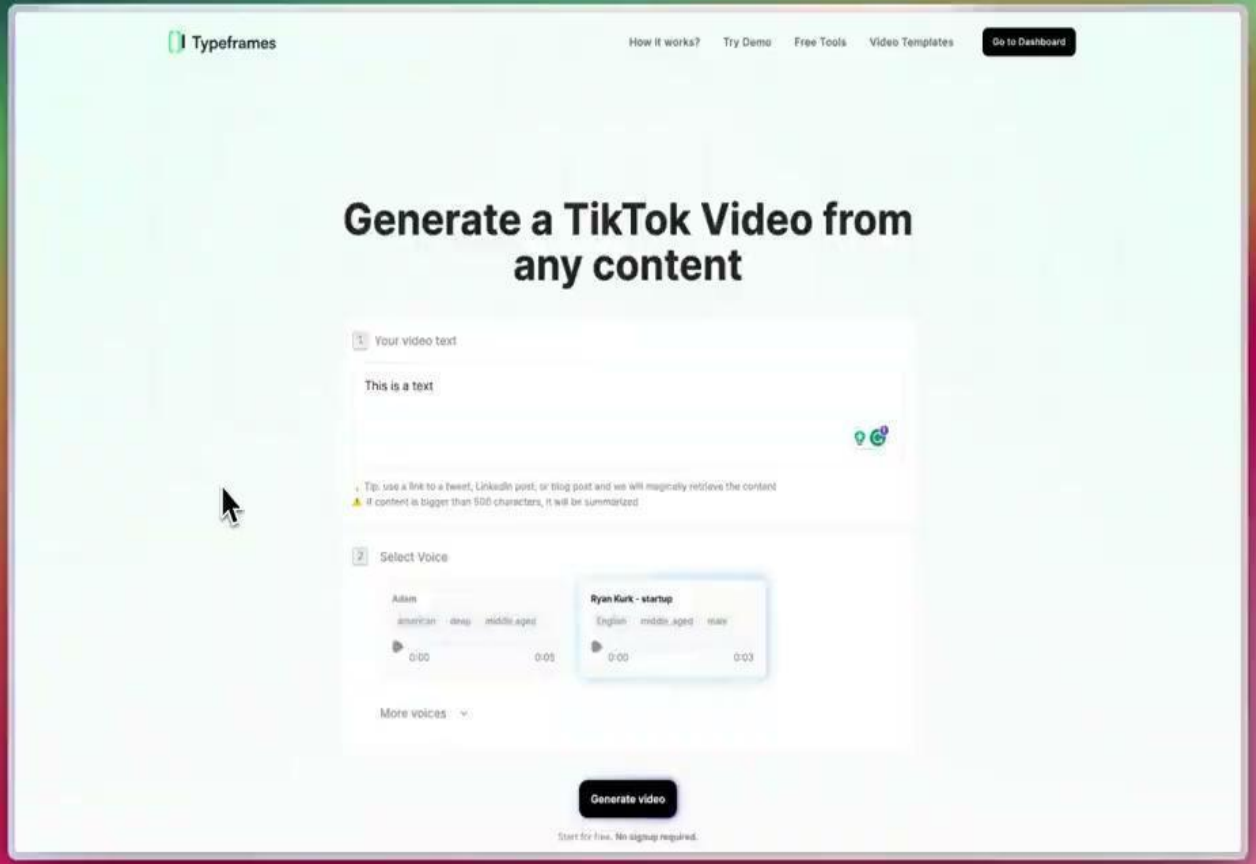


Nano


GOING
THROUGH
TALK TO
AT


Text:
+1 (314) 333-1111
Web:
pi.ai
pi.ai
iOS App:
Pi, your personal AI
Pi
Pi
WhatsApp:
+1 (314) 333-1111
Telegram:
pi.ai/tg
Instagram:
@heypi.ai
Facebook
Messenger


For illustrative purposes only
META
Juan Pino
Research Manager


Typeframes
How it works?
Try Demo
Free Tools
Video Templates
Go to Dashboard
Generate a TikTok Video from any content
1 Your video text
This is a text
Tip: use a link to a tweet, LinkedIn post, or blog post and we will magically retrieve the content
If content is bigger than 500 characters, it will be summarized
2 Select Voice
Adam
american deep middle_aged
0:00
0:05
Ryan Kurk - startup
English middle_aged male
0:00
0:03
More voices
Generate video
Start for free. No signup required.

Reference Image Driving Pose Ours

MagicAnimate:

Temporally Consistent Human Image Animation using Diffusion Model

Zhongcong Xu[1], Jianfeng Zhang[2], Jun Hao Liew[2], Hanshu Yan[2], Jia-Wei Liu[1], Chenxu Zhang[2], Jiashi Feng[2], Mike Zheng Shou[1]

[1]Show Lab, National University of Singapore [2]Bytedance


Original
RESEMBLE.AI


RESEMBLE.AI

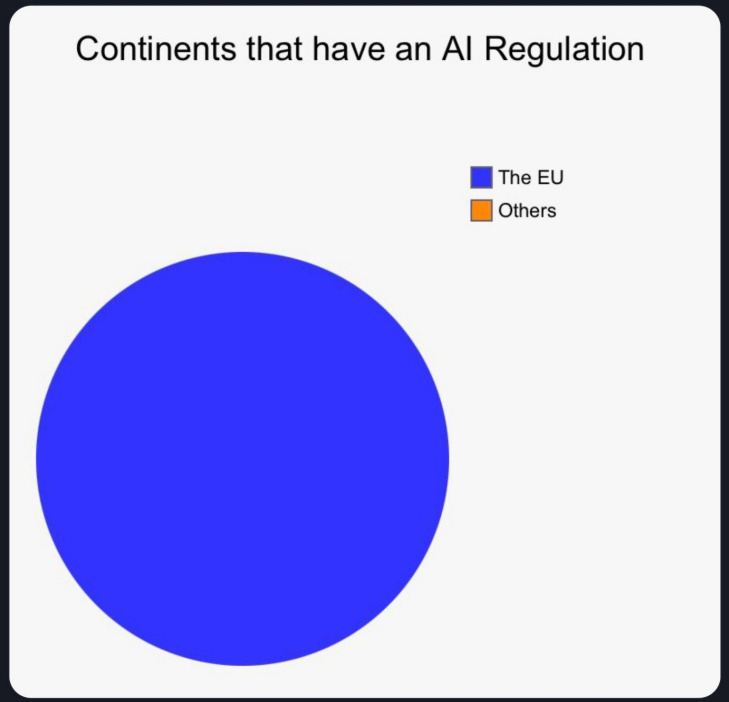

DOMO ROBOT


Musicfy

Thierry Breton
@ThierryBreton

Deal!

#AIAct

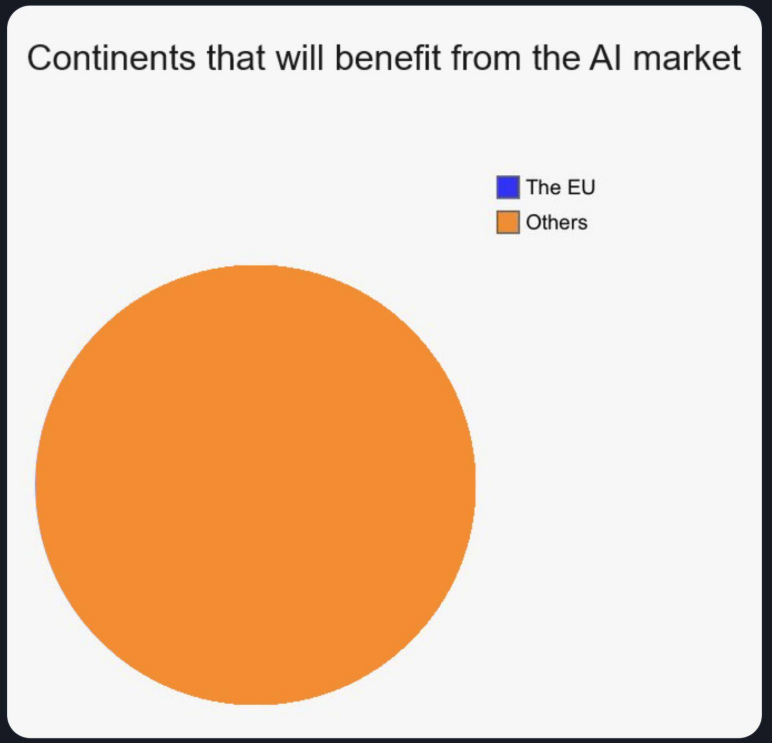


10:37 PM · 8 Dec 2023

1788 3762

DIMENSION
@DIMENSION_YT

@ThierryBreton



2:30 PM · 9 Dec 2023

5 713

#CONF'

"Le RAG : booster de connaissances pour la GenIA ? "

Barbara DELACROIX
& Marvin SANT

Fondateurs @ Devana.ai

devana.ai

Le 19/12/23 à 19h

le wagon