



# *Computer Vision con OpenAI e OpenCV*

Moreno Gentili

**CodicePlastico**

# A story told by

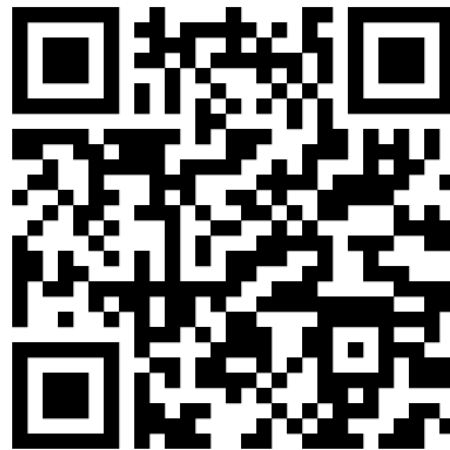


MORENO GENTILI

[moreno.gentili@codiceplastico.com](mailto:moreno.gentili@codiceplastico.com)



CODICEPLASTICO

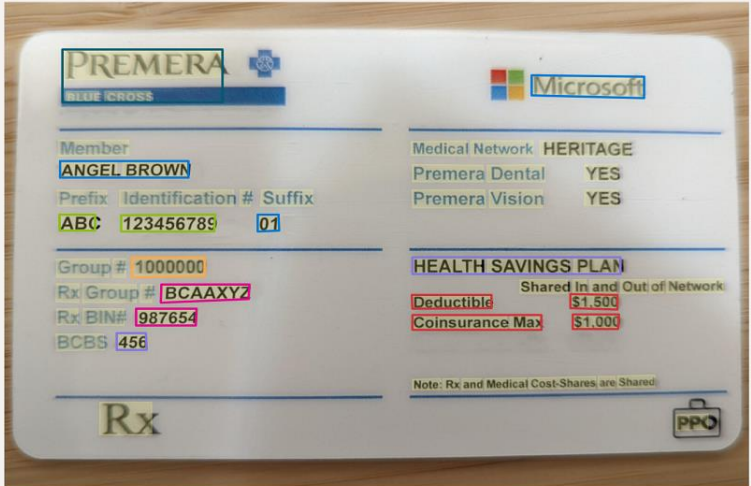


[/Moreno-Gentili/cv-demo](#)



# Azure AI Document Intelligence Studio?

Run analysis | Analyze options



< 1 of 1 >

Fields Result Code

Copays (2) #1

1 Amount  
\$1,500  
Benefit  
Deductible

2 Amount  
\$1,000  
Benefit  
Coinsurance Max

GroupNumber #1 99.50%  
1000000

IdNumber #1 99.50%  
Number 99.50%  
123456789  
Prefix 99.50%  
ABC

Insurer #1 99.50%  
PREMERA BLUE CROSS

# Altre soluzioni?

« [Machine Learning](#)

## Amazon Textract

Estrai automaticamente testo stampato, scrittura a mano, elementi di layout e dati da qualsiasi documento.

[Inizia a usare Amazon Textract](#)

[Contatta l'ufficio commerciale](#)

**L'elaborazione intelligente dei documenti di Amazon offre un ROI del 73%.**

[Compila un breve modulo per scaricare il report.](#)

Favorisci una maggiore efficienza aziendale e un processo decisionale più rapido riducendo al contempo i costi.

Estrai informazioni dettagliate con elevata precisione da qualsiasi documento.

Aumenta o riduci la pipeline di elaborazione dei documenti per adattare rapidamente l'attività alle richieste del mercato.

Automatizza l'elaborazione dei dati in modo sicuro con standard di privacy dei dati, di crittografia e di conformità.



## Altre soluzioni?

<https://openai.com/index/new-models-and-developer-products-announced-at-devday/>

November 6, 2023

# New models and developer products announced at DevDay

GPT-4 Turbo with 128K context and lower prices, the new Assistants API, GPT-4 Turbo with Vision, DALL·E 3 API, and more.



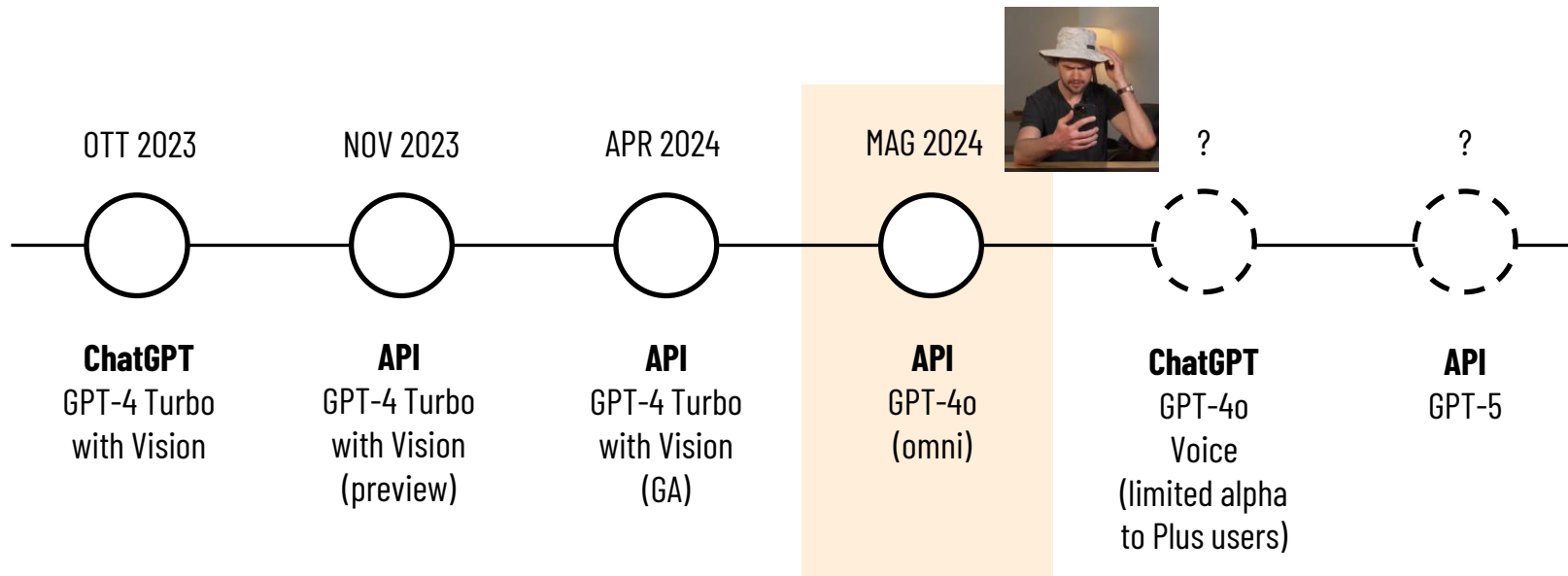
# GPT-4

Capisce il linguaggio naturale

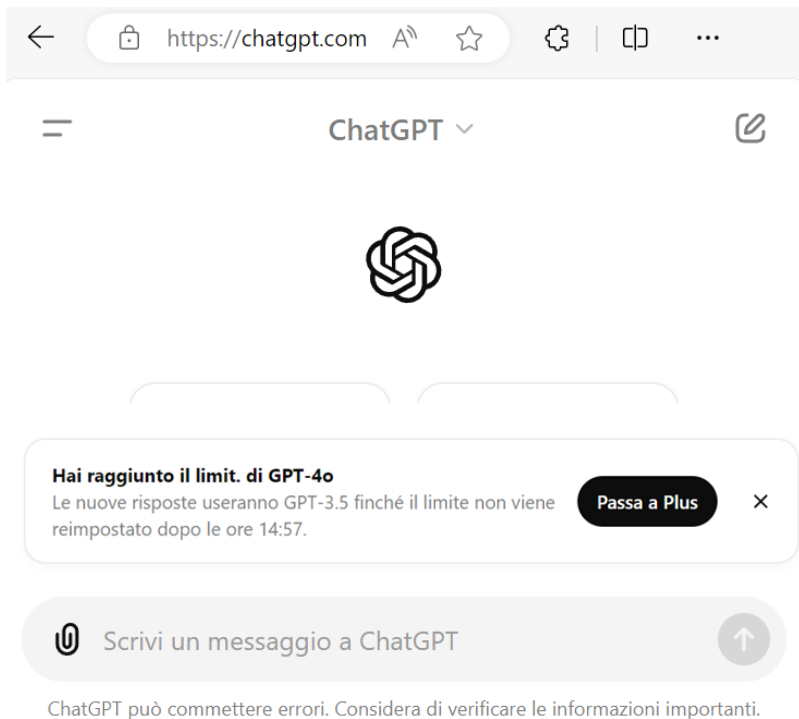
Capacità di ragionamento

Può lavorare con testo e/o immagini

È un modello linguistico di grandi dimensioni multi-modale (MLLM)



# Disponibile via ChatGPT e via API



## Making requests

You can paste the command below into your terminal to run your first API request. Make sure to replace `$OPENAI_API_KEY` with your secret API key. If you are using a legacy user key and you have multiple projects, you will also need to **specify the Project Id**. For improved security, we recommend transitioning to project based keys instead.

```
1 curl https://api.openai.com/v1/chat/completions \
2   -H "Content-Type: application/json" \
3   -H "Authorization: Bearer $OPENAI_API_KEY" \
4   -d '{
5     "model": "gpt-4-turbo",
6     "messages": [{"role": "user", "content": "Say this"}],
7     "temperature": 0.7
8   }'
```



# Ottenere una API Key di OpenAI

- Inizia da <https://platform.openai.com/api-keys>
- Crea un account oppure accedere con il vostro account Google, Microsoft o Apple;
- Fornisci il tuo numero di telefono e avvia la procedura di verifica;

Verify your phone number to create an API key

Start verification





# Ottenere una API Key di OpenAI

- Dopo esserti loggato e aver verificato il numero, crea la API Key

Personal / Demo

Dashboard Docs API reference

API keys

+ Create new secret key

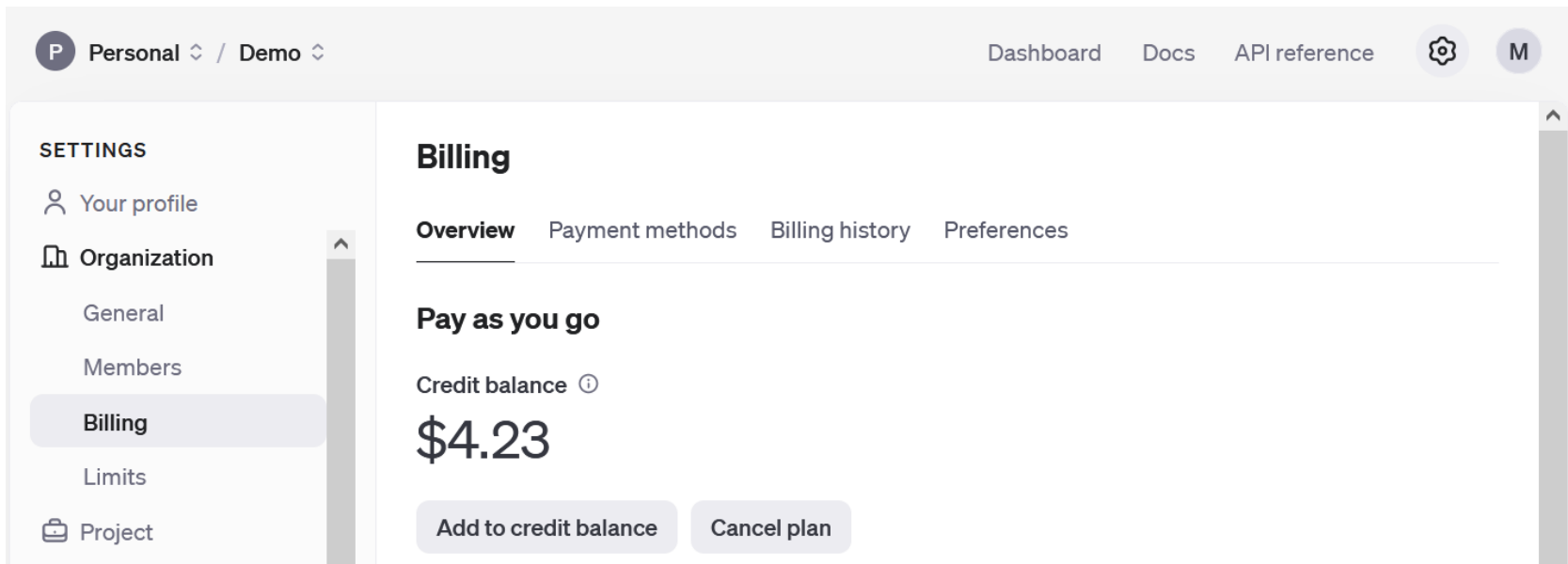
**Project API keys have replaced user API keys.**  
We recommend using project based API keys for more granular control over your resources. [Learn more](#)

View user API keys

NAME	SECRET KEY	CREATED BY	PERMISSIONS
Secret key	sk-...i8pU	demo-account	All

# Ottenere una API Key di OpenAI

- Dalla sezione Billing <https://platform.openai.com/settings/organization/billing/overview>
- Aggiungi almeno \$5.00 di credito.



The screenshot shows the OpenAI platform settings interface. At the top, there's a navigation bar with 'Personal' and 'Demo' dropdowns, and links for 'Dashboard', 'Docs', 'API reference', a settings gear icon, and a user profile 'M'. On the left, a 'SETTINGS' sidebar lists 'Your profile', 'Organization', 'Billing' (highlighted), 'Limits', and 'Project'. The main content area is titled 'Billing' and has tabs for 'Overview', 'Payment methods', 'Billing history', and 'Preferences'. Under the 'Overview' tab, it says 'Pay as you go' and 'Credit balance' with an info icon. The credit balance is displayed as '\$4.23'. At the bottom, there are two buttons: 'Add to credit balance' and 'Cancel plan'.

Personal ▾ / Demo ▾

Dashboard Docs API reference ⚙️ M

**SETTINGS**

- 👤 Your profile
- 🏢 Organization
  - General
  - Members
  - Billing**
  - Limits
- 📁 Project

**Billing**

Overview Payment methods Billing history Preferences

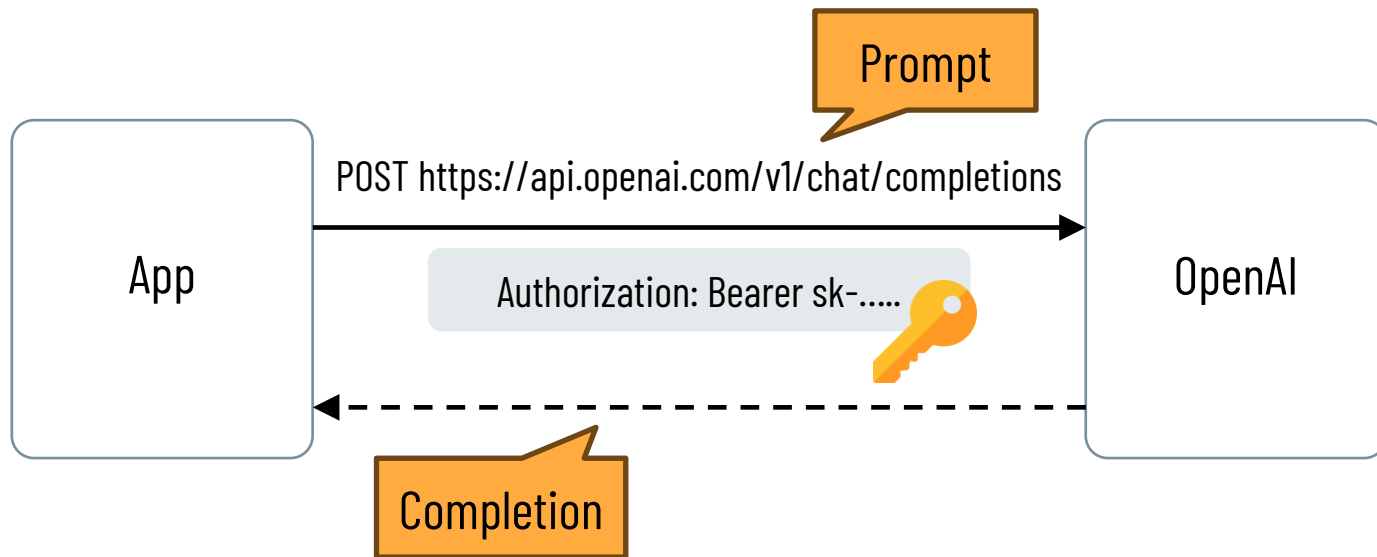
**Pay as you go**

Credit balance ⓘ

**\$4.23**

Add to credit balance Cancel plan

# Usare una API Key di OpenAI





**DEMO**

# Chat Completions API

<https://platform.openai.com/playground/complete>



# Prompt engineering

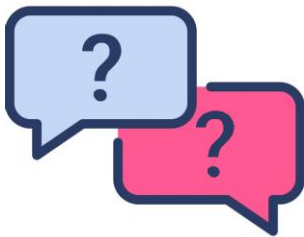
Ottenere migliori risultati scegliendo opportunamente il prompt.

<https://platform.openai.com/docs/guides/prompt-engineering/prompt-engineering>



## Istruzioni chiare

Fornite esempi; attenzione alla punteggiatura.



## Spezzare le domande

Definire chiaramente gli step a cui si vuole risposta.



## Chiedere la verifica

Anziché generare soluzioni, meglio farle verificare.



## Testo di riferimento

Fornite il testo su cui state facendo la domanda.

## Senza testo di riferimento...

Allucinazione!

*Info obsoleta?*

Posso portare il mio gatto al ristorante DrCat di Brescia?



Sì, il ristorante DrCat di Brescia è noto per essere un cat café, quindi i gatti sono i benvenuti. Tuttavia, è sempre una buona idea chiamare in anticipo per confermare la politica attuale del locale riguardo ai gatti ospiti e per verificare se ci sono regole specifiche da seguire.



ChatGPT può commettere errori. Considera di verificare le informazioni importanti.

# OpenAI è ben consapevole delle allucinazioni: CriticGPT

Aiuta gli AI Trainers di OpenAI a identificare le allucinazioni e migliorare il modello.



Task

Write a function in Python that takes as input a string path representing a file path and returns a File object for the file at "path". Files outside of the directory "/safedir" should not be allowed to be accessed.

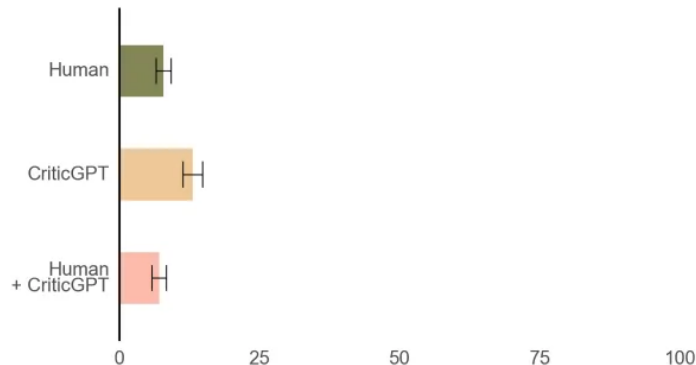


Solution

```
import os

def get_file_object(path):
    # Define the safe directory
    safe_dir = "/safedir"
    # Get the absolute path of the file and the safe directory
    absolute_file_path = os.path.abspath(path)
    absolute_safe_dir = os.path.abspath(safe_dir)

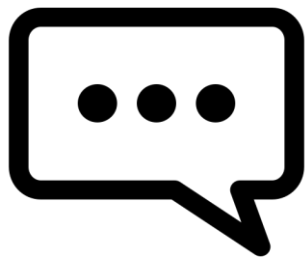
    # Check if the file is within the safe directory
    if not absolute_file_path.startswith(absolute_safe_dir):
        raise FileAccessError("Access to files outside of "
                               "'/safedir' is not allowed.")
    ...
```



CriticGPT

Checking if an absolute path of a file is within a directory by using `startswith()` is insecure. Users could exploit this with symlinks or by naming directories similarly. Using `os.path.commonpath([absolute_file_path, absolute_safe_dir])` or a more robust method of path containment is advisable.

Allucinazioni



## Chat Completions API

### **Interazioni stateless**

Domanda / risposta

Nessuno storico

Basso costo

Con riferimenti, il costo sale

oppure



## Assistants API

### **Interazioni stateful**

Threads / Chat bots

Mantiene il contesto

Costo variabile

Può avvalersi di **file search**

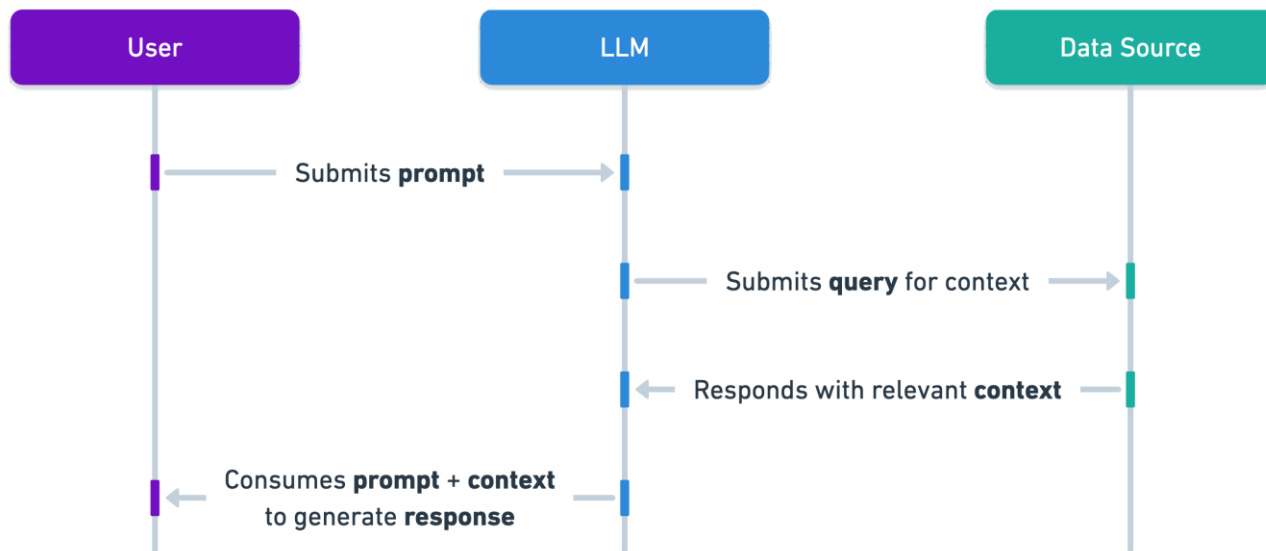
PDF, DOCX, PPTX, ...





# Assistants API: Retrieval Augmented Generation (RAG)

Il modello arricchisce la sua risposta in base al contenuto dei documenti che avremo caricato.



<https://help.openai.com/en/articles/8868588-retrieval-augmented-generation-rag-and-semantic-search-for-gpts>

<https://platform.openai.com/docs/assistants/tools/file-search/supported-files>



**DEMO**

Assistants API

<https://platform.openai.com/playground/assistants>



# Privacy: uso dei contenuti per finalità di training

Il comportamento è diverso a seconda che si tratti di persone fisiche o di imprese.

<https://help.openai.com/en/articles/5722486-how-your-data-is-used-to-improve-model-performance>

PERSONE SU CHATGPT



*When you use our services such as ChatGPT or DALL•E,  
**we may use your content to train our models.***

IMPRESE




***We don't use content from our business offerings  
such as ChatGPT Team, ChatGPT Enterprise, and our  
API Platform to train our models.***




# Privacy: uso dei contenuti per finalità di training


ChatGPT ▾

 ChatGPT Plus  
Il nostro modello più intelligente e molto altro ancora


Fai l'upgrade

 ChatGPT  
Ottimo per le attività quotidiane

✓

 Chat temporanea

☒



## Chat temporanea

Questa chat non sarà visibile nella cronologia e non verrà usata per addestrare i nostri modelli. Per ragioni di sicurezza, potremmo conservarne una copia per un massimo di 30 giorni.



Scrivi un messaggio a ChatGPT



# Quali miei dati sono in possesso di GPT-4?

I servizi di OpenAI sono conformi al GDPR, quindi è possibile vedere i propri dati e chiederne la rimozione.

<https://privacy.openai.com/policies?modal=take-control>



## Download my data

Request a copy of your data



## Do not train on my content

Ask us to stop training on your content



## Delete my OpenAI account

You can ask that we delete your personal data.



## OpenAI Personal Data Removal Request

Remove your personal data from AI model outputs.



*"Ma quanto costa?"*

*"Dipende!"*



# Costi

Dipendono dal modello, dalle dimensioni di prompt e completion.

<https://openai.com/api/pricing/>

	Prompt (input)	Completion (output)
GPT-4o	\$5.00 / 1M token	\$15.00 / 1M token
GPT-4-Turbo	\$10.00 / 1M token	\$30.00 / 1M token
GPT-3.5-Turbo	\$0.50 / 1M token	\$1.50 / 1M token



BATCH API

"Accetto di ricevere il risultato entro 24 ore"

<https://platform.openai.com/docs/guides/batch/getting-started>

# Cos'è un token?

L'unità minima fatturabile, ottenuta spezzando il testo o l'immagine iniziale in parti più piccole.

## Prompt

Alice has a parrot.  
What animal is Alice's pet?

15 token \* \$5.00 / 1M =  
**\$ 0,000075**

## Completion

Alice's pet is a parrot.

8 token \* \$15.00 / 1M =  
**\$ 0,00012**





# Tokenizzazione

Possiamo fare una stima del numero di token con  [/openai/tiktoken](https://github.com/openai/tiktoken)  
Per un'anteprima: <https://platform.openai.com/tokenizer>

```
<?xml version="1.0" encoding="UTF-8"?>
<p:FatturaElettronica versione="FPA12" xmlns:ds="http://www.w3.org/2000/
09/xmldsig#"
xmlns:p="http://ivaservizi.agenziaentrate.gov.it/docs/xsd/fatture/v1.2"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://ivaservizi.agenziaentrate.gov.it/docs/xsd/f
atture/v1.2 http://www.fatturapa.gov.it/export/fatturazione/sdi/fattur
apa/v1.2/Schema_del_file_xml_FatturaPA_versione_1.2.xsd">
  <FatturaElettronicaHeader>
    <DatiTrasmissione>
      <IdTrasmittente>
        <IdPaese>IT</IdPaese>
        <IdCodice>01234567890</IdCodice>
```

Ottimizziamo i costi  
minificando l'XML

*"E le immagini?"*



# Tessellazione

512px

**1 tile = 170 token**

**$4 \text{ tile} * 170 \text{ token} * \$5.00 / 1\text{M} =$**   
**\$0,0034**



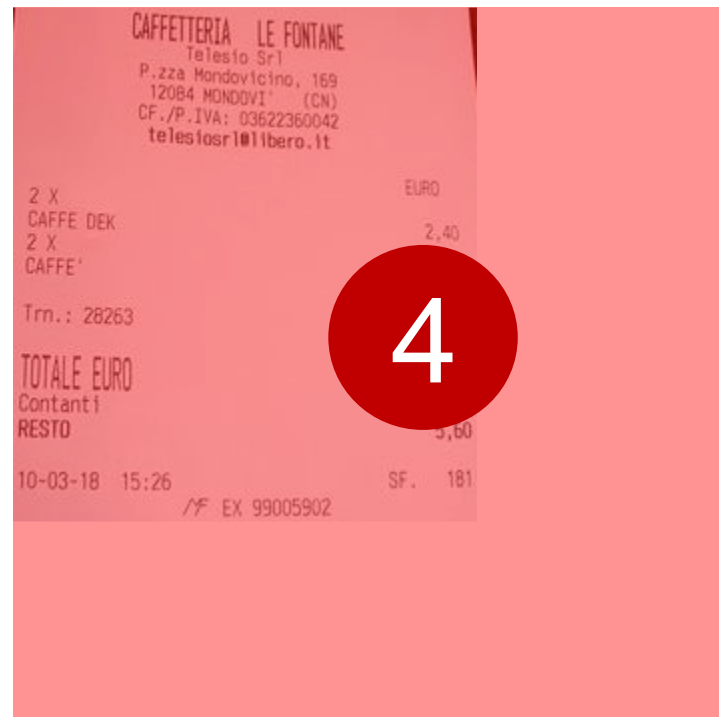
**...ma perché un tile equivale proprio a 170 token?**

<https://www.oranlooney.com/post/gpt-cnn/>



# Ottimizzare i costi delle immagini

512px



# Perché ottimizzare i costi?

- Beh, per spendere meno 😊
- Ma anche per evitare di arrivare al limite imposto da OpenAI.

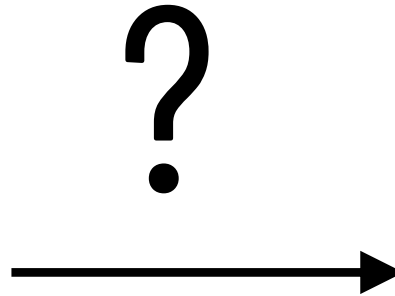
429 Too Many Requests

<https://platform.openai.com/docs/guides/rate-limits/usage-tiers>

Tier	Pagato	Giorni	Richieste / minuto	Token / minuto	Limite coda batch
Tier 1	\$ 5	-	500	30.000	90.000
Tier 2	\$ 50	7	5.000	450.000	1.350.000
Tier 3	\$ 100	7	5.000	600.000	40.000.000
Tier 4	\$ 250	14	10.000	800.000	80.000.000
Tier 5	\$ 1000	30	10.000	2.000.000	300.000.000



*"Come faccio a sapere dove devo tagliare?"*



# Usare modelli di machine learning in locale

Region Based Convolutional Neural Networks (R-CNN)

[https://github.com/berba1995/Faster-RCNN-Receipt-detection/blob/main/work\\_notebook.ipynb](https://github.com/berba1995/Faster-RCNN-Receipt-detection/blob/main/work_notebook.ipynb)



*"Posso assumere che ci sia del testo.  
Taglio attorno al testo."*





# Text detection

Rilevare la presenza di testo in un'immagine.

- Tesseract
- OpenCV con CRAFT Text Detector
- ML.NET con EAST Text Detector

500ms

3000ms

6000ms

A photograph of a menu with red bounding boxes around the text, demonstrating text detection. The menu lists various food and drink items with their prices. The text is slightly blurred and the background is a light blue color.

5	COPERTO	17,50
1	ACQUA PANNA NATURALE	3,50
6	COCA COLA 33CL	9,00
1	CALICE VINO BIANCO	8,00
1	ANTIPASTO MISTO MARINATO	20,00
4	OSTRICHE	16,00
2	FASOLARI	4,00
0,22	GAMBERONI	26,40
0,57	SCAMPI	58,40
0,35	TONNO	22,75
1	POLPO ALLA LUCIANA	20,00
1	INSALATA DI MARE	20,00
1	MIX GRATINATO	17,00
3	MALTAGLIATA ALLO SCOGLIO	58,00
1	PIATTO DEL GIORNO	25,00
1,05	SPAGHETTI ALL'ASTICE	128,00
1	CAFFE AMERICANO	2,50
3	CAFFE	4,50
Servizio (0%):0,00		
Sub Totale:		476,55
Sconto (0,12%):0,55		
Totale:		476,00

*"Troppo lento, troppo intensivo per la CPU"*



# Inoltre...

Tesseract non funziona col testo scritto a mano.



11.12.1995 N. 498

DEBITANTE O PRESTATORE, DIPENDENZA, DOMICILIO O RESIDENZA, CODICE FISCALE, PARTITA IVA

N. ATTA

**MODUS CIBANDI S.p.A.**  
**V. Vincenzo Meli 2/4**  
**Via Pasquale Di Luca 12/24**  
**00152 Roma RM**  
**C.F./P.IVA 12318961001**

DATA  
**03/08/2017**

IDENTIFICATIVI DEL CLIENTE

CODICE FISCALE

PARTITA IVA

QUANTITA	NATURA E QUALITA DEI BENI O DEI SERVIZI	IMPORTO
1	carpa ccid.	13.00
1	risotto.	13.00
1	cous cous.	13.00
1	pecorino.	16.00

RESPONSABILI

Compartecipativo pagato

FATTURA (RICEVUTA FISCALE)

SPINA S.p.A. - Via Marzovilla, 13 - Fiumicino - Autorizzazione Agenzia delle Entrate D.R. Lazio n. 112811/01 del 10-10-2002

*"Non potrei semplicemente tagliare attorno  
alle zone ad alto contrasto?"*



**SİPARİŞ FİŞİ**

**Order**

Tariffs

Mass No.

003552

CİNSİ	MIKTAR	FIYATI	TUTARI
B52	1		13
Cola	2		18
TESTİ ET	1		180
TESTİ ET	1		110
9010			32
			353

**P'AI MIYI' CAI-I-  
RESTAURANT**  
3rd Floor Rm. No : 124A  
Mitsuboshi / ISTANBUL.

**다들 물어봐서**

• • • • •

**1. Learning**

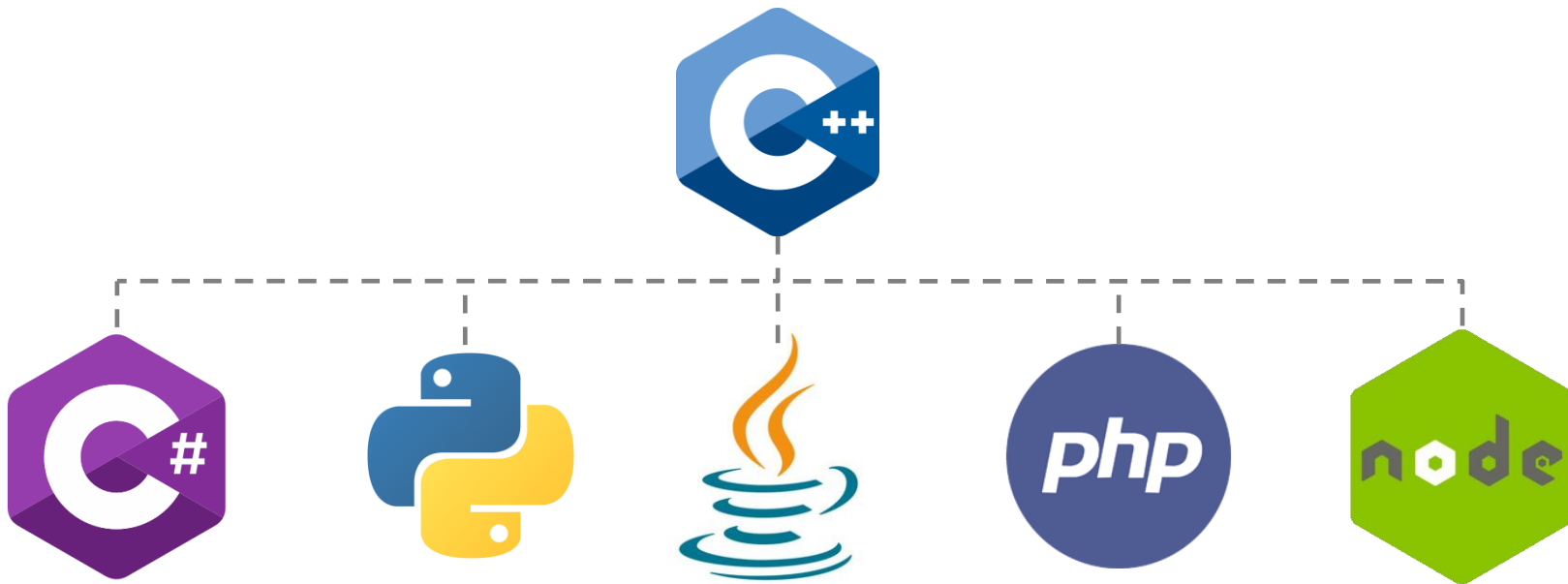
**1980-81**

**0973.2%**

CINAI	MICAR	PICOTI	TSTARI
B-52	✓		17
Coor	20		18
Jest ET	(D)		190
Jost ET	(D)		110
Golo			32
			35?

# OpenCV

È una collezione di 2500 algoritmi di Computer Vision. Originariamente sviluppata da Intel, è una libreria multiplatforma che oggi è gestita dalla OpenCV Foundation. Su licenza Apache 2.0.



# Computer Vision con OpenCV



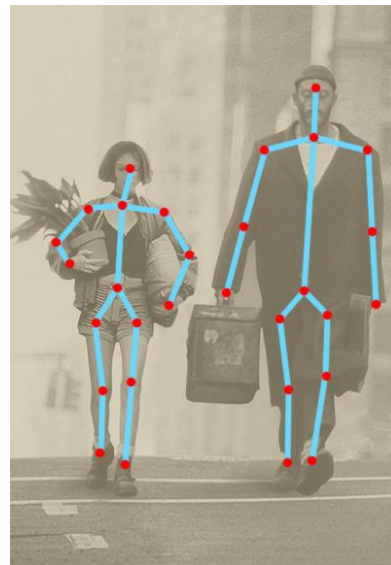
## Elaborazione

Resize, crop, threshold, add, subtract, morph, contour...



## Movimento

Optical flow, background difference su frame seguente



## 3D Point Clouds

Percezione della profondità da camere stereo (es. Kinect)



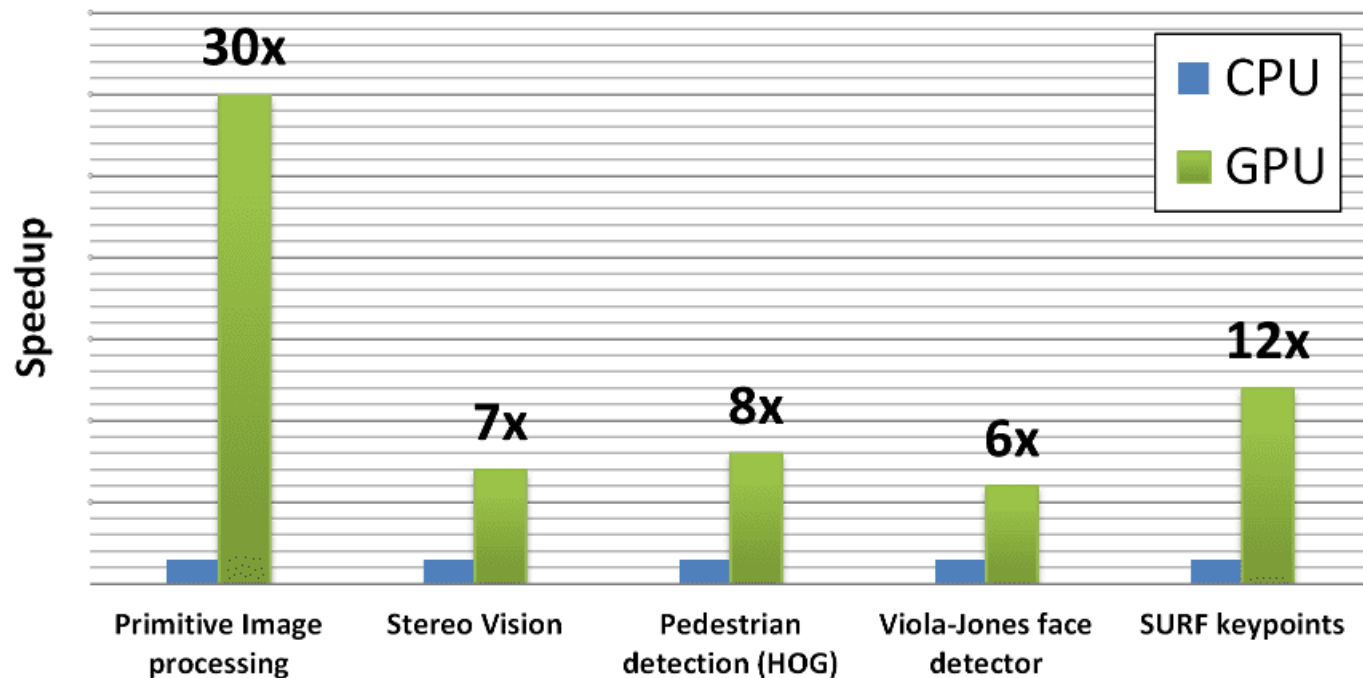
## Detection

Classificazione Haar Cascade, modulo di machine learning



# Computer Vision con OpenCV

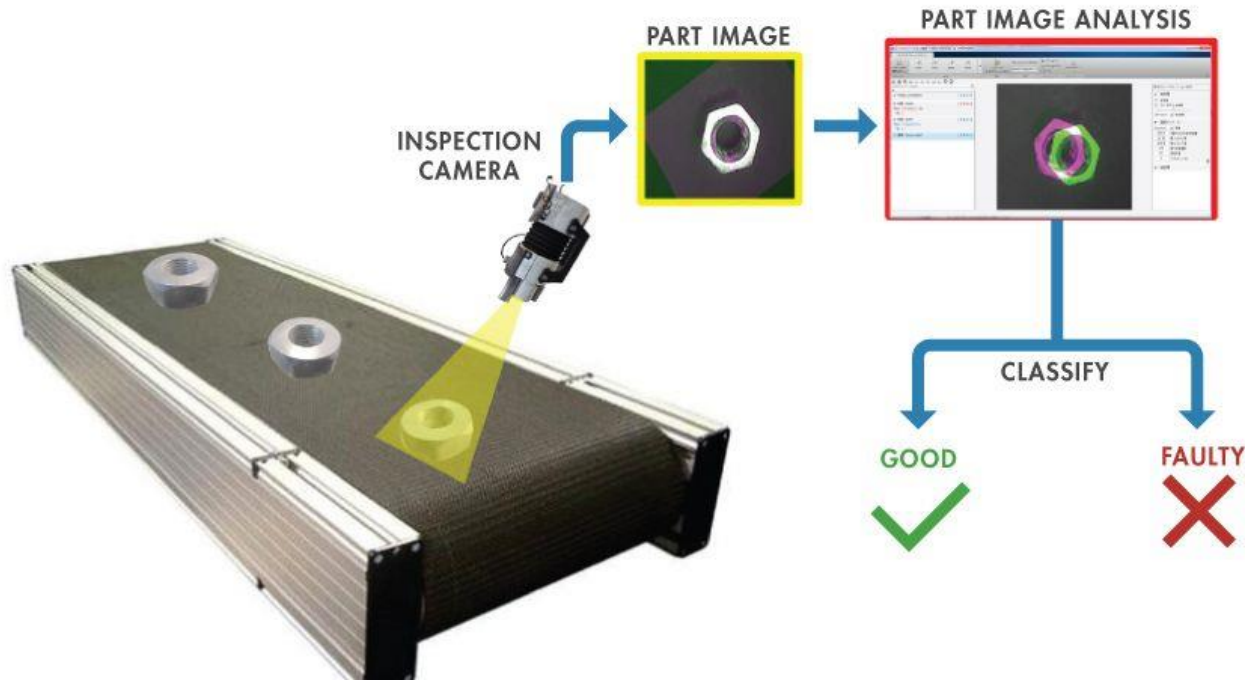
Gli algoritmi possono essere eseguiti su GPU Nvidia (CUDA)





# Computer Vision con OpenCV

Usato nell'industria per la rilevazione dei difetti / controllo qualità.



# Computer Vision con OpenCV

Usato nell'industria per la rilevazione dei difetti / controllo qualità.

Particulates



Weld Lines



Chips



Ideale in situazioni in cui  
l'illuminazione è costante  
(o almeno predicibile)



**DEMO**

Usare OpenCV per rilevare zone ad alto contrasto



# Problemi "black box"

- Li approcciamo senza necessariamente sapere a priori cosa funzioni meglio.
- Sono una **black box**.

☒ AdaptiveThreshold ×

TYPE	BLOCK SIZE		
Binary	7	-	+ → ?

FLAG	C		
GaussianC	20	-	+ → ?

☒ Morph ×

TYPE	SIZE		
Dilate	5	-	+ → ?

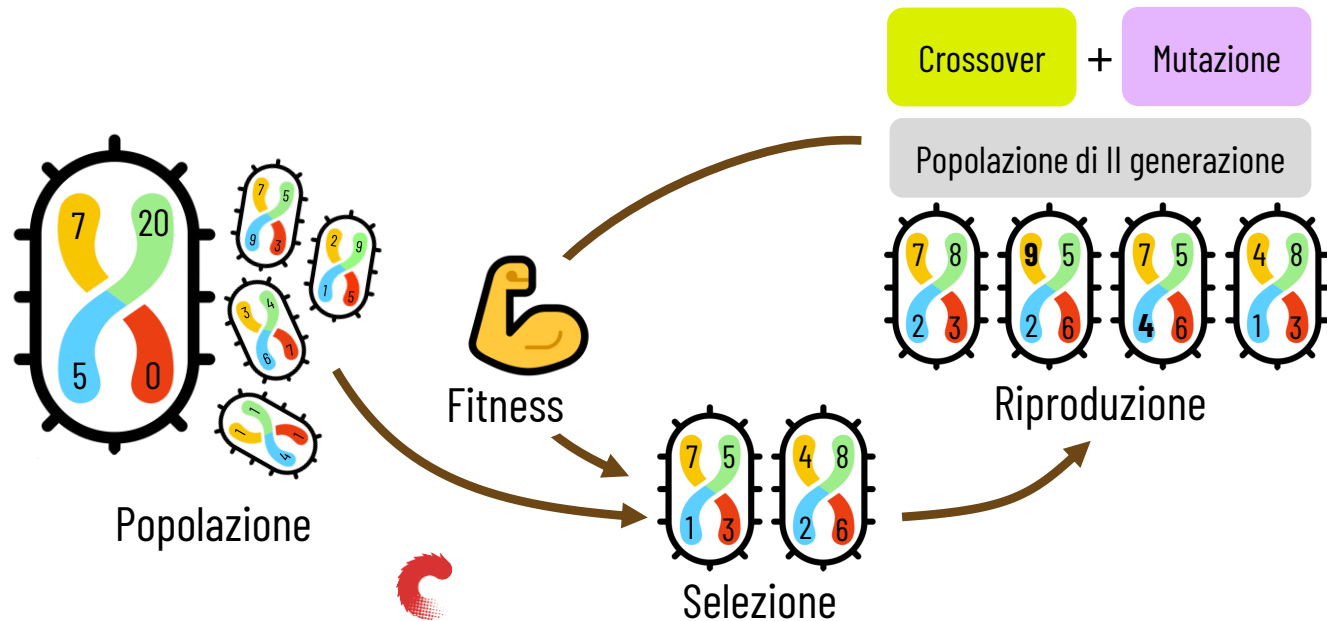
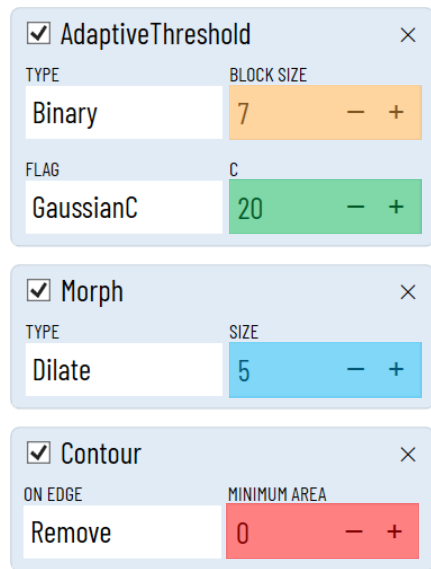
☒ Contour ×

ON EDGE	MINIMUM AREA		
Remove	0	-	+ → ?



# Algoritmi genetici

- Sono algoritmi euristici, ispirati dal processo di selezione naturale;
- I "geni" evolvono nelle generazioni successive finché non diventano utili.



# Determinare la "fitness"

Utile a capire quali individui della popolazione hanno i geni migliori.



Individuo #1

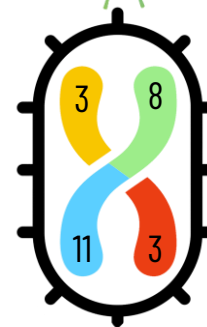
Fitness: 0%



Label

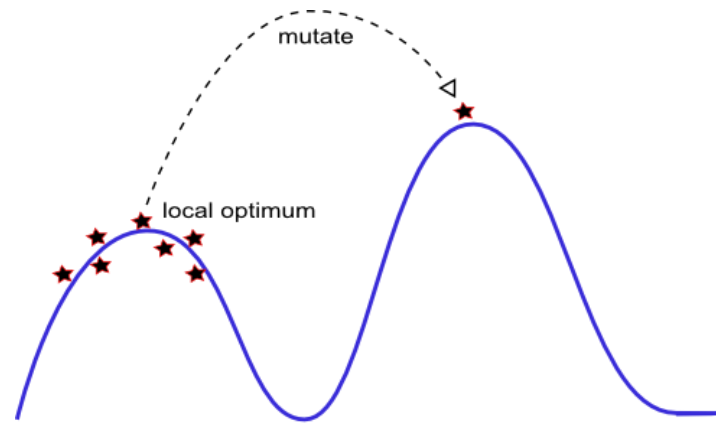
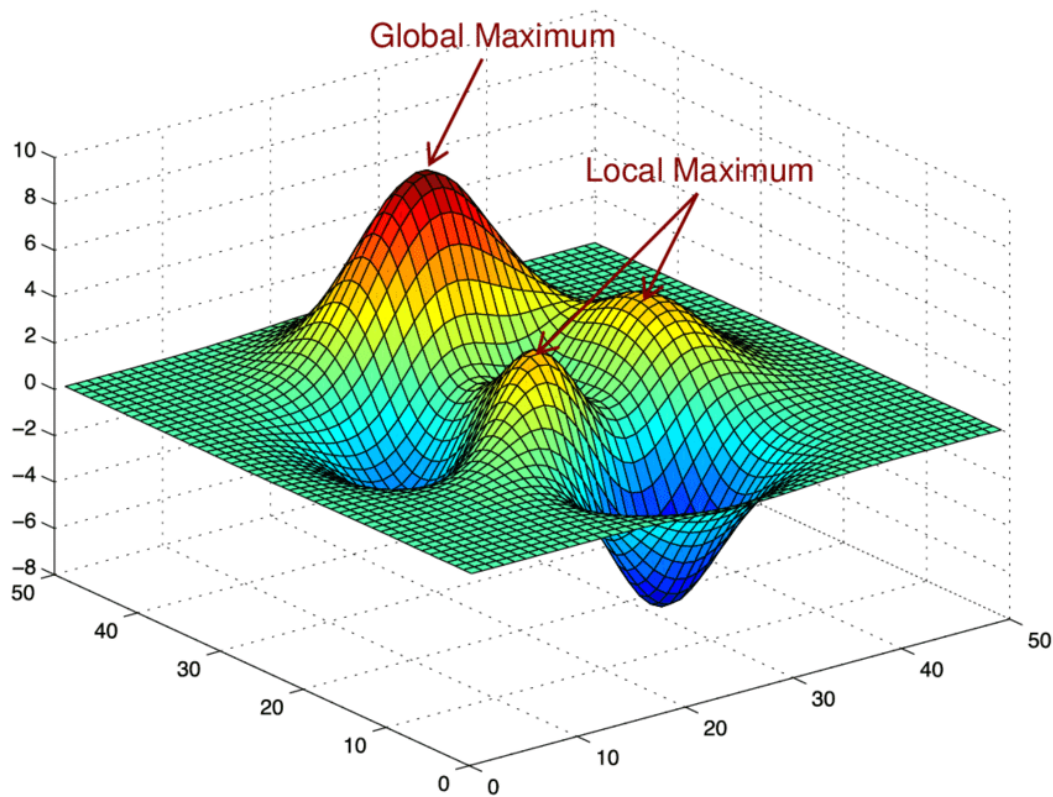
Fitness: 30%

YOU WIN!



Individuo #2

# Le mutazioni sono utili per non incappare negli ottimi locali



## NASA - Space Technology 5 (2006)

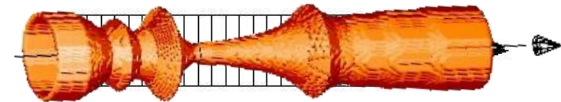
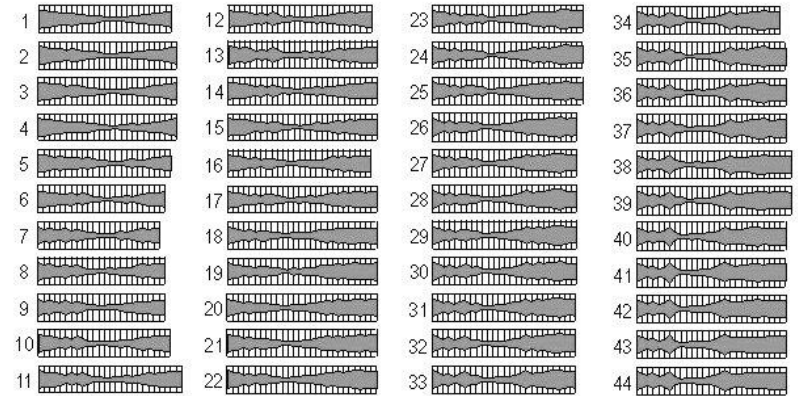
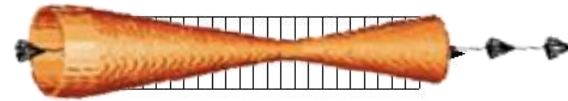
*"Vogliamo migliorare l'antenna dei  
microsatelliti. Come la progettiamo?"*





# Unilever - Ugelli usati nella produzione di detergenti (1960)

*"Vogliamo evitare che gli ugelli si intasino. Come li progettiamo?"*





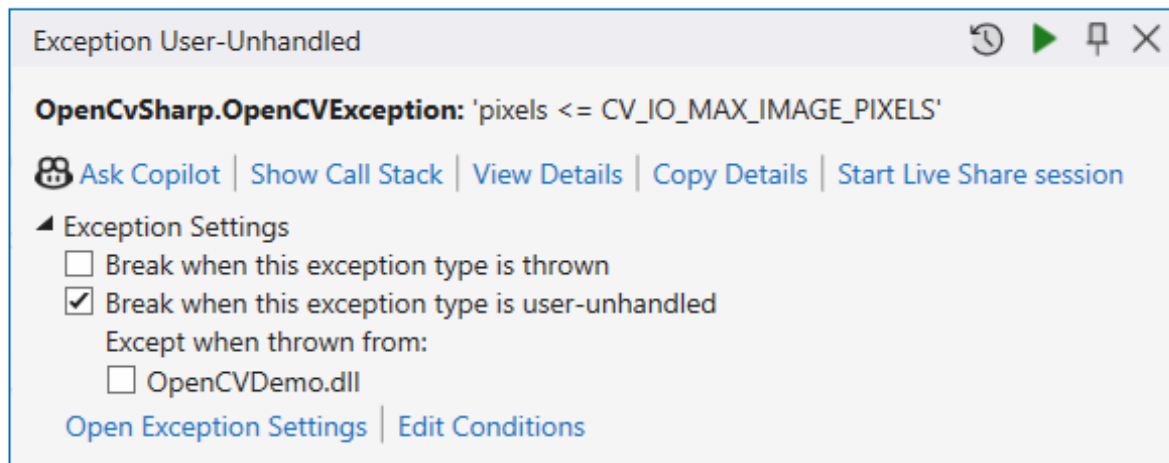
**DEMO**

Usare GeneticSharp trovare i valori migliori



# Impostare la variabile d'ambiente OPENCV\_IO\_MAX\_IMAGE\_PIXELS

Limita il numero di pixel che OpenCV può gestire.



```
Mat originalImage = Cv2.ImDecode(buffer, ImreadModes.Unchanged);
```



# This story was told by



MORENO GENTILI

[moreno.gentili@codiceplastico.com](mailto:moreno.gentili@codiceplastico.com)



CODICEPLASTICO



[/Moreno-Gentili/cv-demo](#)





**CodicePlastico**

FOLLOW US

