

# **Aplicație pentru generare de avataruri AI**

**Autor: David Leonard-Florin, 333AA**

**An: III**

# Cuprins

1. Introducere, prezentare a temei, prezentare a obiectivelor
2. Prezentare pe scurt a suportului ethnic și a realizărilor similare
3. Prezentare tehnică a etapei de realizare/implementare
4. Prezentare mod de utilizare, interacțiune cu utilizatorul, configurare
5. Concluzii
6. Referințe bibliografice

# 1. Introducere, prezentare a temei, prezentare a obiectivelor

## Introducere

În contextul actual, în care inteligența artificială și tehnologiile inovative joacă un rol tot mai semnificativ în viața cotidiană, tema aleasă pentru acest proiect, respectiv generarea automată de avatururi, capătă o importanță deosebită. Am ales această temă deoarece consider că avatarurile digitale sunt tot mai populare și utilizate în diverse medii sociale și profesionale, de la rețele sociale și aplicații de mesagerie până la platforme profesionale și de divertisment.

Relevanța temei este accentuată de tendința tot mai mare a utilizatorilor de a-și personaliza prezența online. Lumea digitală este într-o continuă expansiune, oamenii caută modalități noi și creative de a-și exprima identitatea, personalitatea și preferințele. Astfel, folosirea tehnologiei avansate, precum modele generative AI dezvoltate de OpenAI (precum GPT-4o și DALL·E 3), deschide noi oportunități de interacțiune și personalizare, punând la dispoziția utilizatorului instrumente interactive și intuitive.

## Prezentare a temei

Proiectul propune dezvoltarea unei aplicații web inovatoare care permite utilizatorilor să încarce o simplă fotografie și să obțină rapid un avatar stilizat și personalizat în diferite stiluri artistice: cartoon, anime, Pixar sau 3D. Față de alte proiecte similare existente pe piață, această aplicație aduce un plus major prin utilizarea combinată a tehnologiei avansate de analiză facială (DeepFace) și a modelelor AI dezvoltate de OpenAI, ceea ce permite o analiză mai profundă și mai exactă a caracteristicilor persoanei pentru o generare mai dezvoltată și mai precisă a avatarurilor.

## Prezentare a obiectivelor

Obiectivul principal al aplicației este de a oferi o experiență simplă și plăcută utilizatorului, fără a necesita cunoștințe tehnice aprofundate sau instalări complexe. Aplicația se dorește a fi 100% online și ușor de folosit, oferind astfel acces rapid și facil oricărei persoane interesate să-și personalizeze prezența digitală. Proiectul urmărește să îmbine creativitatea, personalizarea și tehnologia avansată, oferind utilizatorilor o modalitate inovatoare și atractivă de a se exprima în mediul digital actual.

- **Dezvoltarea unei aplicații web interactive** care permite generarea automată a unui avatar pe baza unei imagini reale încărcate de utilizator.
- **Integrarea modelelor AI oferite de OpenAI (GPT-4o + DALL·E 3)** pentru procesarea imaginii și generarea creativă a avatarului în stiluri variate.
- **Utilizarea bibliotecii DeepFace** pentru a extrage trăsături faciale esențiale (vârstă, sex, etnie, expresie) în vederea construirii unui prompt cât mai precis.

- **Implementarea a patru stiluri vizuale de generare a avatarului:** cartoon, anime, Pixar și 3D, pentru a oferi utilizatorului flexibilitate și opțiuni personalizate.
- **Crearea unei interfețe moderne și intuitive,** cu animații vizuale, captură foto prin cameră și feedback vizual în timp real pentru o experiență fluidă.

## 2. Prezentare pe scurt a suportului tehnic și a realizărilor similare

### 2.1. Analiza unor realizări similare existente

În prezent, există numeroase aplicații și platforme care utilizează inteligența artificială pentru generarea avatarurilor și prelucrarea imaginilor, câteva dintre cele mai relevante și cunoscute fiind Bitmoji, FaceApp și platformele bazate pe Midjourney sau Stable Diffusion.

- **Bitmoji** reprezintă o aplicație care permite utilizatorilor să creeze avataruri stilizate, cartoon, ce pot fi integrate în diverse rețele sociale. Punctul forte al Bitmoji îl reprezintă varietatea și personalizarea ridicată a avatarurilor. Totuși, limitarea principală constă în stilul unic, care nu permite diversificarea avatarurilor în alte stiluri vizuale cum ar fi anime sau Pixar.
- **FaceApp** folosește inteligența artificială pentru modificarea facială realistă, oferind efecte diverse cum ar fi schimbarea vârstei, genului sau a expresiilor faciale. Avantajul major este realismul transformărilor, însă punctul slab constă în lipsa unor stiluri artistice specifice, precum cartoon sau 3D.
- **Midjourney și Stable Diffusion** sunt platforme de generare de imagini bazate pe text (prompt-uri), foarte populare în zona artistică și creativă. Ele permit generarea unor imagini foarte detaliate și artistice în diverse stiluri. Deși au capacități puternice, dezavantajul principal este complexitatea creării prompt-urilor, care necesită o cunoaștere avansată a tehnicilor specifice, nefiind foarte accesibile utilizatorilor începători.

Prin urmare, obiectivul principal al aplicației este să combinăm accesibilitatea aplicațiilor precum Bitmoji cu diversitatea stilistică și calitatea oferită de platforme precum DALL·E și GPT-4 de la OpenAI. Aplicația propusă este orientată către utilizatorii obișnuiți care nu au experiență tehnică, dar care doresc să beneficieze de rezultate artistice de calitate superioară, într-o manieră rapidă și simplă.

### 2.2. Aspecte multimedia și fundamentare teoretică

Proiectul nostru se bazează pe tehnologii avansate din sfera multimedia și inteligenței artificiale. Mai jos sunt prezentate detaliat principalele concepte și tehnologii utilizate:

## Generarea imaginilor folosind OpenAI (DALL·E 3)

DALL·E 3 este un model generativ avansat creat de OpenAI, specializat în generarea imaginilor pe baza unor descrieri textuale detaliate numite „prompt-uri”. Principiul fundamental din spatele acestui model este reprezentat de rețelele neuronale convoluționale și de tehnicile avansate de învățare profundă.

Procesul concret presupune:

- **Prompt-ul textual:** Aplicația noastră generează automat un prompt detaliat în funcție de analiza imaginii originale și stilul ales de utilizator (cartoon, anime, Pixar, 3D).
- **Interpretarea textului:** Modelul DALL·E interpretează prompt-ul, extrage elementele vizuale descrise și creează reprezentări interne abstracte ale acestora.
- **Generarea imaginii:** Pe baza reprezentării interne, modelul produce o imagine nouă, stilizată conform specificațiilor furnizate.

Prompt-urile sunt create într-un mod automatizat, integrând detalii extrase cu ajutorul analizei avansate folosind alte modele AI, precum DeepFace și GPT-4o.

## Utilizarea DeepFace în aplicație

DeepFace este un instrument software avansat pentru analiza facială, utilizat pentru extragerea automată a caracteristicilor detaliate din imagini. În cadrul aplicației noastre, DeepFace este folosit pentru a determina automat:

- **Vârsta aproximativă**
- **Genul persoanei**
- **Emoția predominantă**
- **Etnia estimată**
- **Caracteristici faciale suplimentare** (precum păr facial)

Aceste informații sunt fundamentale pentru crearea unui prompt detaliat și realist pentru generarea avatarului de către DALL·E.

## Comunicare și integrarea funcțiilor în backend

În cadrul backend-ului Flask, am realizat integrarea și comunicarea eficientă între componentele software utilizate:

- Imaginea încărcată de utilizator este trimisă către DeepFace pentru analiză.
- Rezultatele analizei DeepFace sunt preluate și utilizate într-un prompt secundar trimis către GPT-4o pentru a descrie în detaliu hainele, accesoriile și fundalul imaginii originale.
- Prompt-ul rezultat de la GPT-4o și datele extrase de DeepFace sunt combinate într-un prompt unic și finalizat pentru generarea avatarului cu ajutorul modelului DALL·E 3.
- Imaginea generată de DALL·E este apoi descărcată, salvată și pusă la dispoziția utilizatorului.

Astfel, comunicarea între funcții este strict secvențială și bine organizată pentru a garanta o generare rapidă și eficientă a avatarurilor personalizate.

## Reprezentarea, compresia și standardele multimedia

Aplicația utilizează reprezentări standard ale imaginilor în format PNG și JPEG. Formatul PNG este ales pentru imaginile generate datorită calității superioare și lipsei

pierderilor la compresie, fiind ideal pentru grafica digitală și ilustrații. JPEG este folosit pentru imaginile inițiale încărcate de utilizator datorită eficienței în raport cu dimensiunea fișierului.

Compresia imaginilor în aceste formate se realizează implicit de bibliotecile software integrate, iar aplicația asigură transferuri eficiente și stocarea compactă a fișierelor rezultate. Aplicația noastră nu utilizează în mod direct procesarea audio sau video, însă cadrul tehnologic și multimedia pe care îl utilizează poate fi extins cu ușurință în viitor pentru a integra și astfel de funcționalități.

### Reprezentarea imaginilor

Imaginile sunt manipulate în formatele standard **JPEG** și **PNG**. Inițial, fișierul încărcat este citit ca **bytes** și convertit în **base64** pentru a fi transmis către OpenAI în format `data:image/jpeg;base64,....`. Această reprezentare permite încorporarea unei imagini într-un mesaj JSON, în lipsa unui URL direct.

### Compresie imagine

Imaginile sunt compresate implicit în timpul prelucrării. DALL·E 3 generează imagini la rezoluție 1024x1024 px în format PNG, cu o compresie fără pierderi. Pentru display și descărcare, nu se alterează calitatea.

### Sunet și video

Aplicația nu utilizează sunet sau procesare video avansată. Totuși, **folosirea camerei web** prin `navigator.mediaDevices.getUserMedia()` implică **streaming video temporar**. Acesta este procesat local (în browser) și transformat într-o imagine PNG cu ajutorul unui canvas HTML5.

## 2.3 Protocoale și comunicare client-server

Aplicația utilizează **protocolul HTTP/HTTPS** pentru comunicarea dintre frontend (React) și backend (Flask Python). Sunt implementate:

- **metoda POST**: utilizată pentru a trimite imaginea și stilul selectat către server, în endpoint-ul `/upload_and_generate_avatar`.
- **metoda GET**: utilizată pentru a descărca avatarul generat, prin endpoint-ul `/download/<filename>` și pentru a accesa imaginea generată la `/processed/<filename>`.

Transferul fișierelor se realizează printr-un obiect de tip `FormData`, care asigură compatibilitatea cu uploadul imaginilor și trimiterea metadatelor.

## 2.4 Comunicare între componente

Fluxul tehnic este:

1. Utilizatorul încarcă o imagine sau face una cu camera → este salvată local în browser și trimisă către backend via POST.
2. Backend-ul:
  - folosește DeepFace pentru analiză facială
  - trimite imaginea la GPT-4o pentru descriere
  - construiește promptul final

- apelează DALL·E 3 pentru a genera avatarul.
- 3. Răspunsul include un URL spre imaginea generată, servită local din directorul /processed.

### 3. Prezentare tehnică a etapei de realizare/implementare

Realizarea aplicației a presupus parcurgerea mai multor etape tehnice, fiecare implicând decizii legate de tehnologiile utilizate, configurări, integrarea modelelor AI și dezvoltarea front-end-ului. În cele ce urmează sunt prezentate toate componentele principale care au contribuit la realizarea completă a platformei „Avatar Generator AI”.

#### 3.1 Framework-uri, biblioteci și API-uri utilizate

Proiectul utilizează următoarele framework-uri și biblioteci:

- **React.js** – pentru interfața utilizatorului, datorită modularității, performanței și suportului extins pentru componentizare.
- **Flask (Python)** – pentru partea de backend, fiind un microframework ușor de configurat și eficient pentru aplicații REST.
- **OpenAI API** – pentru generarea prompt-urilor și imaginilor cu GPT-4o și DALL·E 3.
- **DeepFace** – pentru analiza imaginii și extragerea trăsăturilor faciale (sex, vârstă, emoție, etnie).
- **TensorFlow & Keras (prin DeepFace)** – pentru funcționarea modelelor de recunoaștere facială.
- **HTML5 Media API** – pentru utilizarea camerei web și capturarea imaginilor în browser.
- **Bootstrap & CSS personalizat** – pentru stilizarea aplicației.

#### 3.2 Structura funcțională

Aplicația este compusă din două componente principale:

- **Frontend (React):** Permite încărcarea imaginii sau capturarea ei cu camera, alegerea stilului avatarului, trimiterea cererii la server și afișarea rezultatului generat.
- **Backend (Flask + AI):** Primește imaginea și stilul, procesează imaginea cu DeepFace, generează un prompt cu GPT-4o și creează un avatar stilizat cu DALL·E 3.

#### 3.3 Algoritm și flux de lucru

Fluxul principal de lucru este următorul:

1. Utilizatorul încarcă o imagine sau o capturează cu camera.
2. Alege unul dintre cele 4 stiluri: Cartoon, Anime, Pixar, 3D.
3. Imaginea este transmisă backend-ului printr-un request POST către endpoint-ul Flask /upload\_and\_generate\_avatar.
4. Backend-ul:
  - Salvează temporar imaginea.
  - Rulează **DeepFace** pentru a extrage trăsături faciale.

- Trimite imaginea și trăsăturile către **GPT-4o** pentru descriere detaliată.
  - Construiește un prompt final pentru **DALL·E 3**.
  - Salvează imaginea generată și trimite un răspuns JSON cu linkul către avatarul final.
5. Frontend-ul afișează imaginea generată și oferă opțiunea de descărcare.

### 3.4 Contribuția personală

Acest proiect a fost realizat integral, fără folosirea template-urilor, cu implicare directă în toate componentele:

- **Frontend:**
  - Implementarea interfeței React de la zero.
  - Integrarea camerei prin HTML5 Media API.
  - Adăugarea efectelor vizuale (fade-in, preview live).
  - Controlul stării aplicației (loading, preview, erori).
- **Backend:**
  - Integrarea cu API-ul OpenAI pentru GPT-4o și DALL·E.
  - Procesarea imaginilor cu DeepFace.
  - Combinarea trăsăturilor detectate cu descrierea generată automat.
  - Gestionarea imaginilor temporare și salvarea rezultatelor.
- **Prompt Engineering:**
  - Adaptarea prompt-urilor pentru fiecare stil (cartoon, anime, pixar, 3D) pe baza unor exemple vizuale.
  - Îmbinarea trăsăturilor detectate cu descrierea GPT pentru rezultate cât mai fidele imaginii inițiale.

### 3.5 Aspecte legate de Aplicații Multimedia (AM)

Aplicația se bazează puternic pe procesarea imaginilor și generarea de conținut multimedia:

- **Analiză facială automată** – cu modele AI pre-antrenate pe seturi mari de date.
- **Generare de imagini sintetice** – folosind modele de difuzie și rețele neuronale autoregresive (DALL·E 3).
- **Procesare video (camera)** – capturarea în timp real a unei imagini folosind camera laptopului.
- **Transformare format imagine** – convertirea de la .jpg sau captură video în .png.
- **UX multimedia** – cu elemente vizuale animate, previzualizări și feedback în timp real pentru utilizator.



## 4. Prezentare mod de utilizare, interacțiune cu utilizatorul, configurare

Aplicația „Avatar Generator AI” a fost dezvoltată pentru a oferi o experiență intuitivă, fluidă și prietenoasă, chiar și pentru utilizatori fără experiență tehnică. În această secțiune vom detalia modalitatea de instalare și rulare, interacțiunea cu aplicația, comportamentul fiecărui element de interfață și opțiunile oferite.

### 4.1 Configurare și rulare

#### Backend (Flask + Python)

1. **Instalarea dependențelor:**

Se rulează comanda: `pip install -r requirements.txt` Acest fișier include toate bibliotecile necesare: Flask, OpenAI, deepface, python-dotenv, etc.

2. **Setarea cheii OpenAI:**

În fișierul `.env`, trebuie adăugată cheia personală: `OPENAI_API_KEY=sk-xxxxxxxxxxxxxxxxxxxxxxxx`

3. **Pornirea serverului:**

Se rulează: `python app.py`

Serverul va porni pe adresa `http://127.0.0.1:5000`.

#### Frontend (React.js)

1. **Navigarea în directorul aplicației:** `cd frontend`

2. **Instalarea modulelor Node:** `npm install`

3. **Pornirea aplicației:** `npm start`

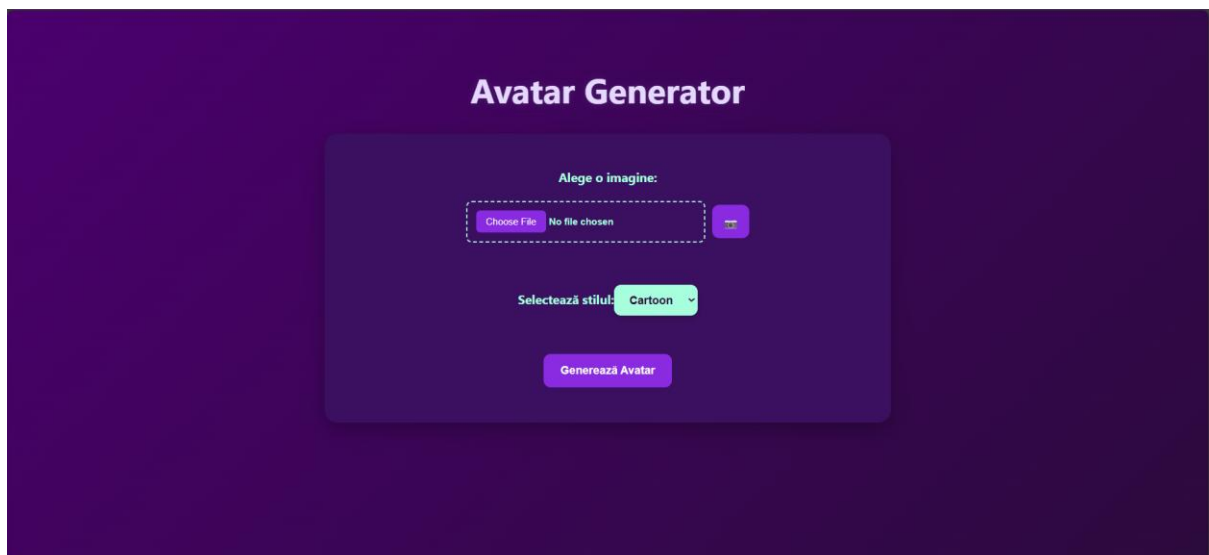
Aplicația va porni implicit pe `http://localhost:3000`.

### 4.2 Interacțiunea utilizatorului

Interfața aplicației este gândită să ghideze utilizatorul pas cu pas în procesul de creare a avatarului:

#### Pasul 1 – Alegerea imaginii

- Utilizatorul poate **încărca o imagine** din calculator folosind butonul de tip „Choose File”.
- Alternativ, poate **folosi camera laptopului** printr-un buton. Acesta activează camera video și permite capturarea unei imagini PNG în timp real.
- Imediat după selectare sau captură, se afișează **o previzualizare a imaginii** pentru verificare.

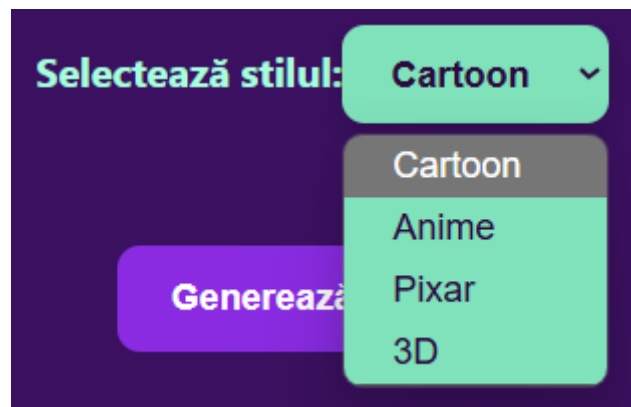


Imagine 1 – Pagina de generare a avatarurilor

### Pasul 2 – Alegerea stilului

- Se poate selecta unul din cele 4 stiluri:
  - **Cartoon** – desen clar cu contururi accentuate.
  - **Anime** – trăsături specifice animației japoneze.
  - **Pixar** – stil 3D prietenos, inspirat de animațiile Pixar.
  - **3D** – redare realistă tip „flat 3D” modern.

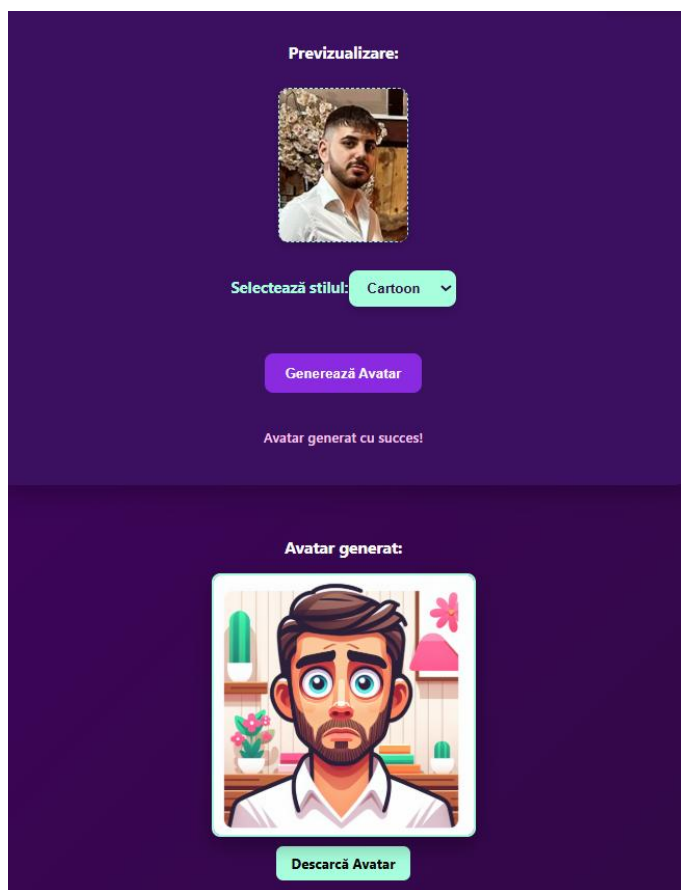
Stilul selectat influențează prompt-ul trimis către DALL·E.



Imagine 2 – Alegerea stilului avatarului

### Pasul 3 – Generarea avatarului

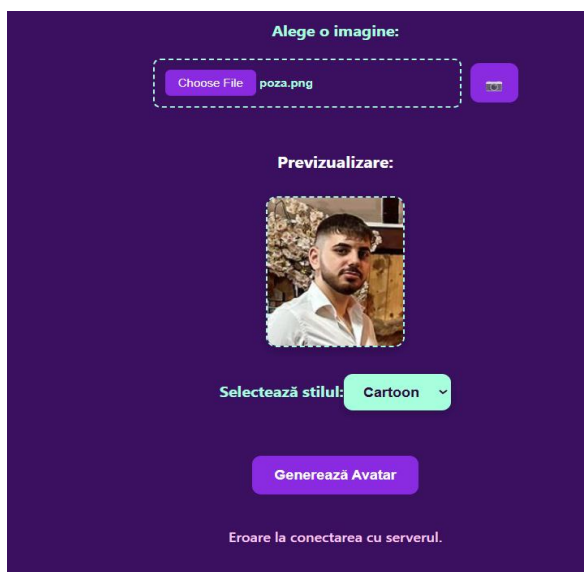
- Prin apăsarea butonului „Generează Avatar”, imaginea este transmisă către backend.
- Un **loader animat** este afișat pe durata procesării.
- La final, este afișat avatarul generat și un buton de **descărcare a imaginii**.



Imagine 4 – Generarea efectivă a unui avatar cu stilul selectat si poză aleasă

#### 4.3 Funcționalități și feedback dinamic

- Aplicația oferă feedback textual în timp real: mesaje de eroare, mesaje de succes, status-ul procesării.
- Toate secțiunile (previzualizare, avatar generat) apar progresiv, cu animații tip fade-in.
- Dacă utilizatorul încercă să genereze un avatar fără imagine, este informat clar despre lipsa fișierului.



Imagine 4 – Afișare feedback dinamic pentru utilizator

#### 4.4 Elemente multimedia și UX

- Aplicația integrează direct **captura video în browser**, o funcționalitate modernă și specifică aplicațiilor multimedia.
- Previzualizarea live a avatarului și a imaginii inițiale oferă un plus de **interactivitate**.
- Designul a fost gândit pentru a evidenția clar pașii de urmat, folosind culori complementare, layout echilibrat și butoane intuitive.

### 5. Concluzii

Proiectul „Avatar Generator AI” a avut ca scop realizarea unei aplicații web interactive care permite utilizatorului să genereze automat un avatar stilizat, pornind de la o imagine încărcată sau capturată în timp real. Din punct de vedere al obiectivelor propuse, toate au fost atinse cu succes:

- **Interfață interactivă complet funcțională:** Aplicația oferă un mod simplu, intuitiv și modern de interacțiune, permițând încărcarea unei imagini sau folosirea camerei pentru a genera un avatar personalizat.
- **Generarea automată a prompt-urilor prin analiză AI:** Imaginea este analizată cu ajutorul DeepFace, iar datele extrase sunt combinate cu descrierea generată de GPT-4o, rezultând un prompt coerent, creativ și contextual.
- **Integrarea unui sistem complet de generare imagini AI (DALL·E):** Folosind OpenAI DALL·E 3, aplicația reușește să producă imagini cu stiluri vizuale variate (cartoon, anime, Pixar, 3D), cu rezultate de calitate superioară.
- **Previzualizare live, animații și descărcare finală:** Utilizatorul primește feedback vizual și textual în fiecare etapă, de la previzualizarea imaginii până la obținerea și descărcarea avatarului generat.
- **Respectarea standardelor de aplicații multimedia:** Proiectul se bazează pe elemente esențiale ale domeniului AM: captură video, generare și procesare imagine, reprezentare digitală, compresie, interfață vizuală animată.

#### Utilitate și actualitate

Într-un context digital în care avatarele personalizate devin din ce în ce mai importante pentru rețele sociale, aplicații de mesagerie, gaming sau metavers, un astfel de instrument vine ca o soluție extrem de **actuală și utilă**.

Majoritatea aplicațiilor existente se bazează fie pe alegeri manuale (forme de ochi, nas, păr etc.), fie pe modele de deep learning antrenate fix. În schimb, acest proiect:

- integrează **mai multe modele AI** (vizual, textual, generativ),
- adaptează promptul automat pentru o **stilizare mai inteligentă și precisă**,
- permite **captura în timp real** a imaginii, oferind un grad de libertate superior.

### Performanță și relevanță

Din punct de vedere tehnic, aplicația rulează fluent în medii locale și poate fi adaptată ușor pentru producție. Modelul backend asigură o procesare eficientă, iar frontend-ul React oferă o interfață fluidă și reactivă.

În termeni de performanță:

- timpul mediu de generare a unui avatar este de **10–15 secunde**,
- avatarul generat are o **rezoluție mare (1024x1024 px)**,
- prompturile sunt adaptate stilului și trăsăturilor detectate, ceea ce duce la **variație ridicată** și rezultate **vizual coerente**.

### Avantaje față de alte aplicații

Față de aplicațiile similare precum Bitmoji, ReadyPlayerMe sau Zmoji, care oferă rezultate fie prea stilizate, fie limitate de opțiuni presetate, „Avatar Generator AI”:

- oferă o **combinație unică între procesare vizuală + generare textuală + randare AI**;
- are **capacitate de învățare contextuală**, folosind descrierea automată a pozei în promptul artistic;
- permite personalizare prin alegerea stilului și oferă **feedback în timp real**, lucru rar întâlnit.

## 6. Referințe bibliografice

### Lucrări științifice și surse tehnice:

1. Ian Goodfellow, Yoshua Bengio, Aaron Courville – **Deep Learning**, MIT Press, 2016, ISBN: 9780262035613  
→ Carte de bază în înțelegerea rețelelor neuronale, utilizată ca fundament pentru componentele de analiză vizuală și generare AI.
2. O'Reilly Media – **Flask Web Development (2nd Edition)**, Miguel Grinberg, 2018, ISBN: 9781491991732  
→ Utilizată pentru dezvoltarea și structurarea arhitecturii backend cu Flask.
3. Serengil, S., & Ozpinar, A. – **LightFace: A Hybrid Deep Face Recognition Framework**, 2020  
DOI: 10.48550/arXiv.2007.06521  
→ Lucrare care stă la baza DeepFace, utilizat pentru analiza fețelor în aplicație.

---

### Referințe web/documentație oficială:

4. OpenAI – *DALL·E 3 Documentation*  
Link: <https://platform.openai.com/docs/guides/images>  
Ultima accesare: 6 mai 2025  
→ Documentație utilizată pentru integrarea modelului DALL·E 3 în backend.

5. DeepFace – *Face Recognition and Attribute Analysis*  
Link: <https://github.com/serengil/deepface>  
Ultima accesare: 6 mai 2025  
→ Bibliotecă folosită pentru extragerea automată a trăsăturilor din imagine (sex, vârstă, etnie, expresie facială).
6. OpenAI – *GPT-4o API Documentation*  
Link: <https://platform.openai.com/docs/guides/gpt>  
Ultima accesare: 6 mai 2025  
→ Folosit pentru generarea automată a descrierii textuale și traducerea acesteia.