

# SZABADON HOZZÁFÉRHETŐ ARCFELISMERÉSI KÖNYVTÁRAK ADATVÉDELMI SZEMPONTÚ VIZSGÁLATA

CSIKTUSNÁDI-KISS KENÉZ – BSC  
SZAKDOLGOZAT

KONZULENSEK: DR. GULYÁS GÁBOR GYÖRGY ÉS FÁBIÁN ISTVÁN



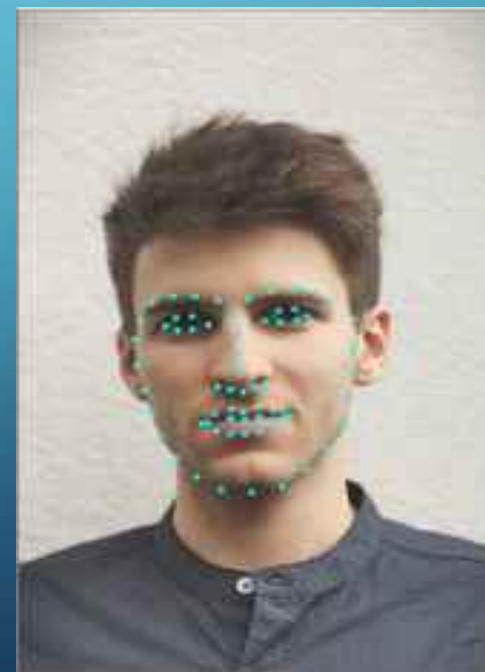
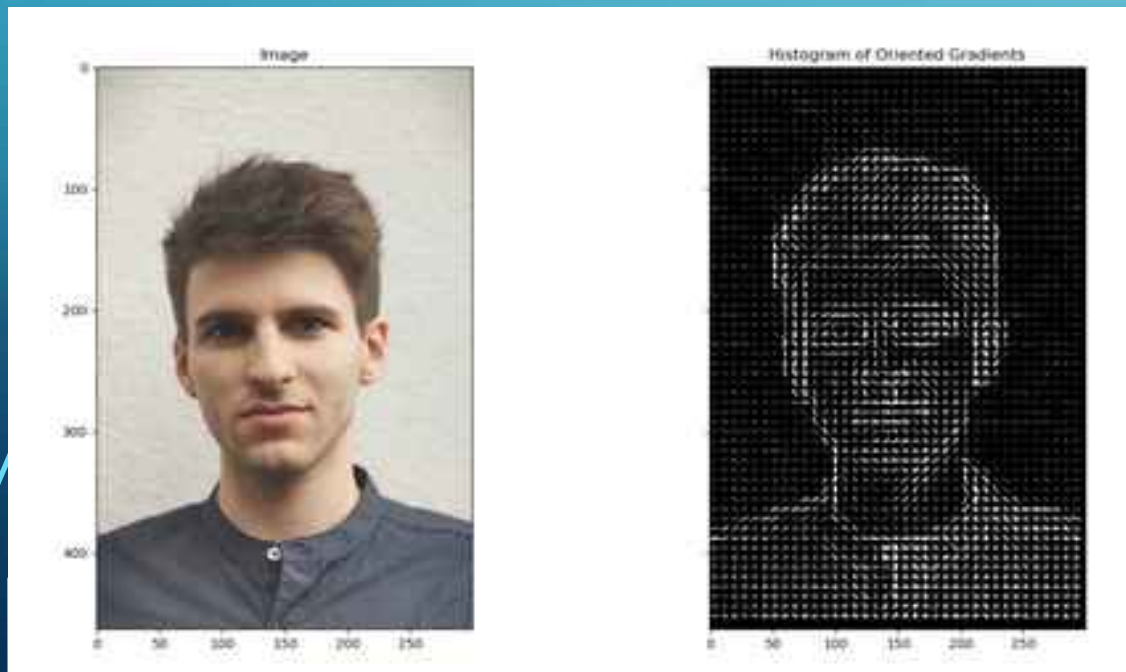
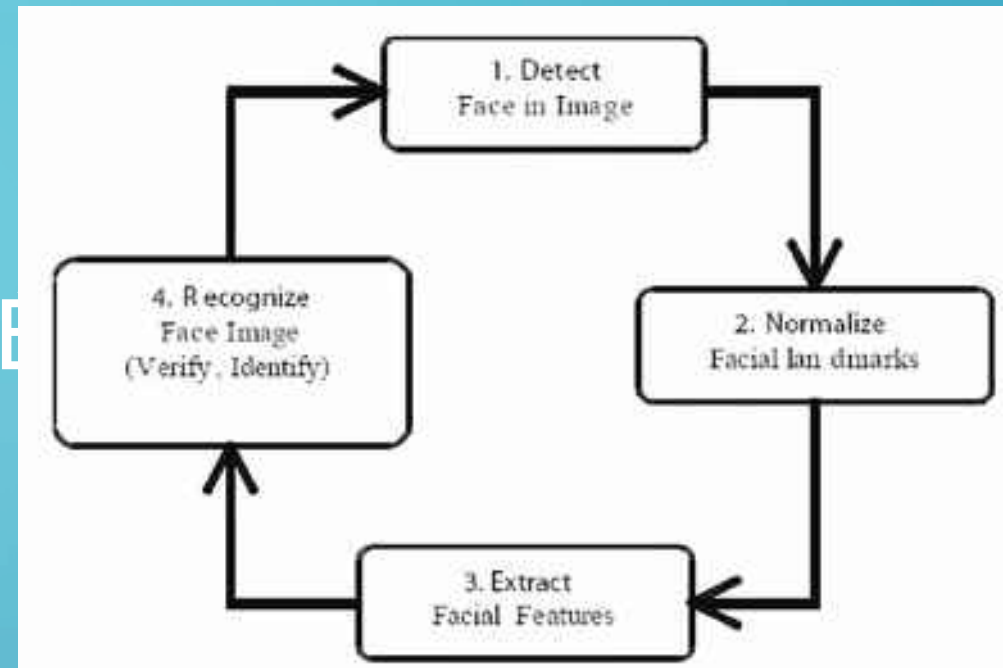
**Budapesti Műszaki és Gazdaságtudományi  
Egyetem**

Villamosmérnöki és Informatikai Kar  
Automatizálási és Alkalmazott Informatikai  
Tanszék

# A MUNKÁM CÉLJA

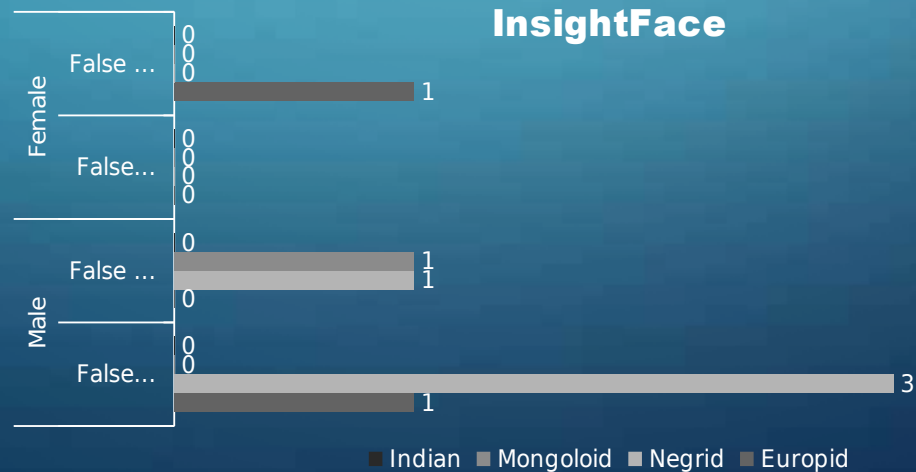
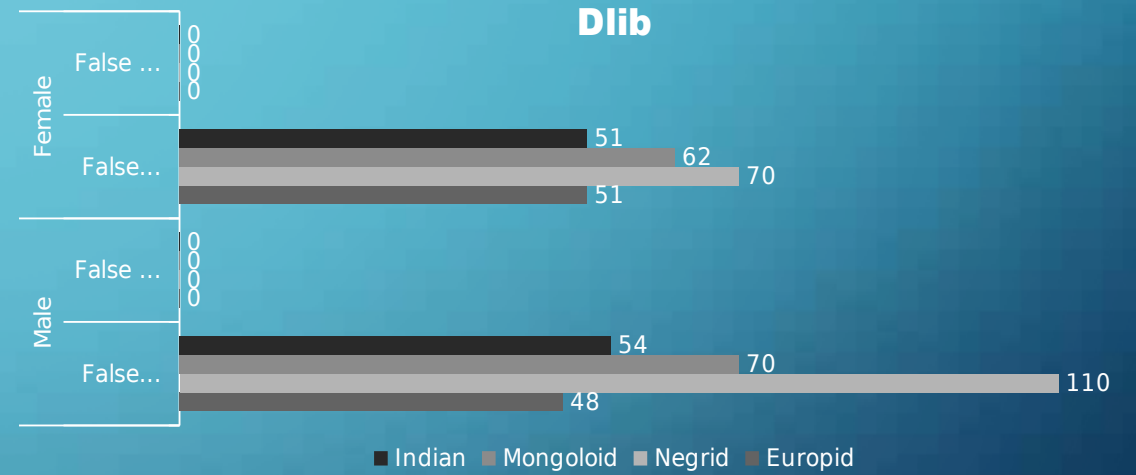
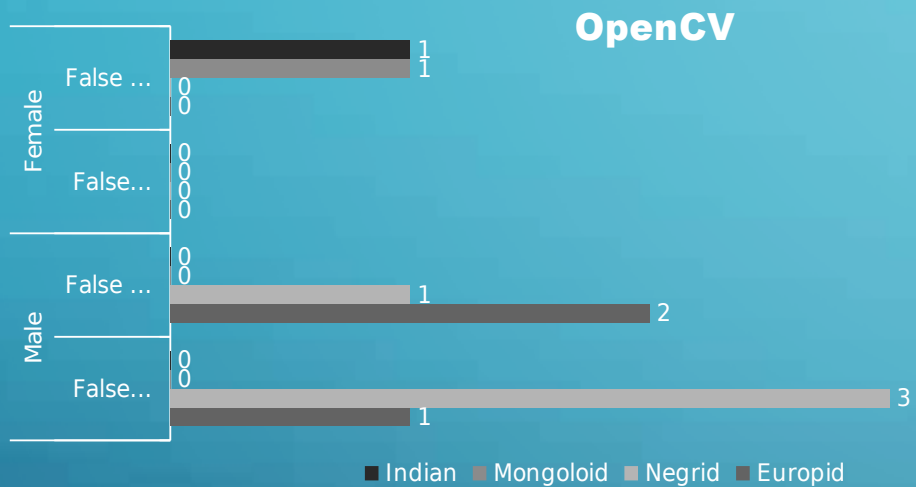
- Elemezni három nyílt forráskódú arcfelismerési eljárást működését
- Összehasonlítani ezen könyvtárak működése során létrejövő ún. embedding-eket, aszerint, hogy milyen mértékben lehet belőlük demográfiai adatokat visszaállítani
- Megvizsgálni, hogy az arcfelismerési folyamat során kinyert biometrikus adatok, milyen adatbiztonsági kockázatokat hordoznak

# AZ ARCFELISMERÉS MŰKÖDÉSE

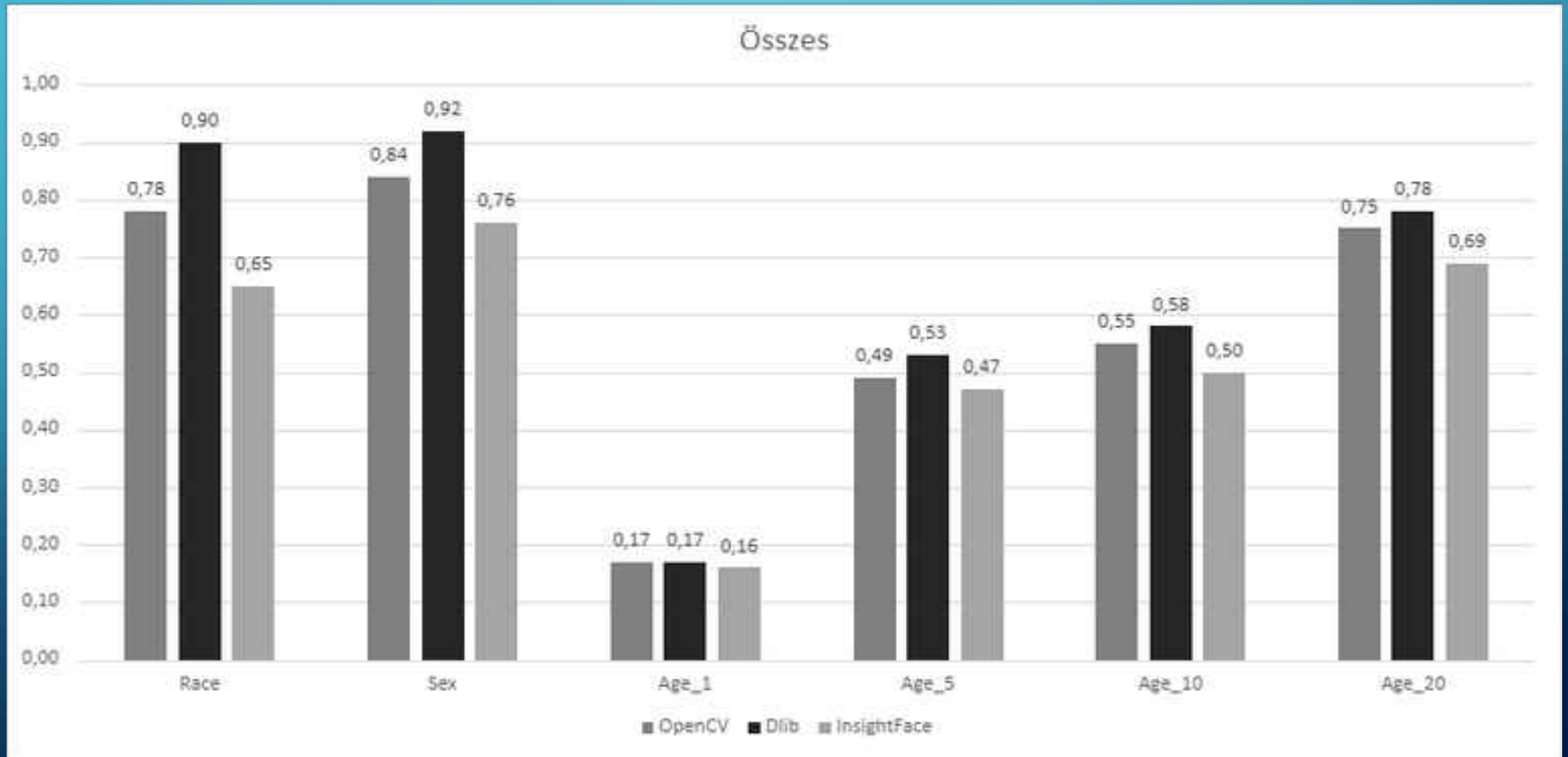


0.341300799916608	0.28140003911201676	0.014395791208111776	-0.05368821308175149
0.08346003015203476	0.014395791208111776	-0.05368821308175149	-0.18278870913181309
0.014395791208111776	-0.05368821308175149	-0.18278870913181309	0.06411513719134801
-0.05368821308175149	-0.18278870913181309	0.06411513719134801	0.01483967080752132
-0.18278870913181309	0.06411513719134801	0.01483967080752132	0.02773474156856517
0.06411513719134801	0.01483967080752132	0.02773474156856517	0.14922991368884915
0.01483967080752132	0.02773474156856517	0.14922991368884915	-0.0688780553153990
0.02773474156856517	0.14922991368884915	-0.0688780553153990	0.1893882918418181
0.14922991368884915	-0.0688780553153990	0.1893882918418181	0.00798951481809148
-0.0688780553153990	0.1893882918418181	0.00798951481809148	-0.2942348887380121
0.1893882918418181	0.00798951481809148	-0.2942348887380121	0.0107960171154785
0.00798951481809148	-0.2942348887380121	0.0107960171154785	-0.01780189757752418
-0.2942348887380121	0.0107960171154785	-0.01780189757752418	0.0502642616629485
0.0107960171154785	0.0502642616629485	0.0502642616629485	-0.0019888833652588
0.0502642616629485	-0.0019888833652588	-0.0019888833652588	0.0228111751148196
-0.0019888833652588	0.0228111751148196	0.0228111751148196	-0.1878894981615879
0.0228111751148196	-0.1878894981615879	-0.1878894981615879	-0.1886346125682722
0.1878894981615879	-0.1886346125682722	-0.1886346125682722	0.01821581888881785
-0.1886346125682722	0.01821581888881785	0.01821581888881785	0.01885736857470
0.01821581888881785	0.01885736857470	0.01885736857470	-0.068227758996411
0.01885736857470	-0.068227758996411	-0.068227758996411	0.0051287858721175
-0.068227758996411	0.0051287858721175	0.0051287858721175	-0.11806888208174
0.0051287858721175	-0.11806888208174	-0.11806888208174	-0.11806888208174
-0.11806888208174	-0.11806888208174	-0.11806888208174	0.05130889913414181
-0.11806888208174	0.05130889913414181	0.05130889913414181	-0.166812420168681
0.05130889913414181	-0.166812420168681	-0.166812420168681	-0.01051151495923988
-0.166812420168681	-0.01051151495923988	-0.01051151495923988	-0.2878489177578661
0.01051151495923988	-0.2878489177578661	-0.2878489177578661	0.00789320275878
-0.2878489177578661	0.00789320275878	0.00789320275878	0.001184497665264187
0.00789320275878	0.001184497665264187	0.001184497665264187	-0.11806888208174
0.001184497665264187	-0.11806888208174	-0.11806888208174	0.05130889913414181
-0.11806888208174	0.05130889913414181	0.05130889913414181	-0.166812420168681
0.05130889913414181	-0.166812420168681	-0.166812420168681	-0.01051151495923988
-0.166812420168681	-0.01051151495923988	-0.01051151495923988	-0.2878489177578661
0.01051151495923988	-0.2878489177578661	-0.2878489177578661	0.00789320275878
-0.2878489177578661	0.00789320275878	0.00789320275878	0.001184497665264187

# KÖNYVTÁRAK STATISZTIKÁI



# EMBEDDINGEK STATISZTIKÁI



# A TÁMADÓ MODELL BEMUTATÁSA

Kiinduló  
adat

- Embedding amit deanonimizálni akarunk

Prediktor  
ok

- A demográfiai adatok kinyerése

Közösség  
i média

- Azonos demográfiai adatokkal rendelkező profilok keresése

Lehetség  
es  
találatok

- Keresett Embeddinggel való összevetés

Megtalált  
profil

- A kiinduló Embeddinghez tartozó személy

# TÁMADÁS ELLENŐRZÖTT KÖRNYEZETBEN

Célja, hogy statisztikai adatokkal szolgáljon arról, hogy a támadási modell, milyen hatékonysággal képes egy embedding alapján egy személyt azonosítani

- A siker ráta: **36.7%**

```
match_candidates = {}
for i in range(0, len(deanonim_candidates_df)):
    target = deanonim_candidates_df.loc[[i], :]
    if (int(target.at[i, 'Sex']) == int(sex_prediction) and
        int(target.at[i, 'Race']) == int(race_prediction) and
        (
            int(target.at[i, 'Age']) - 20 <= int(age_prediction) <= int(target.at[i, 'Age']) + 20
        )):
        target = target.drop(columns=['Sex', 'Race', 'Age'])
        result = face_recognition.compare_faces([target.to_numpy().astype(np.float)[0]],
                                                embedding_to_deanonim.to_numpy().astype(np.float)[0], tolerance=0.5)
        if result[0]:
            match_candidates[i] = target.to_numpy().astype(np.float)[0]
```

# TÁMADÁS A KÖZÖSSÉGI MÉDIA BEVONÁSÁVAL

A közösségi médiában a keresési szempontok:

- A személyek halmazát kellő mértékben szűkítsék egy sikeres azonosításhoz
- Az általuk megtalált személyeknek legyen olyan közös kapcsolódási pontjuk, mely miatt elképzelhető, hogy az arcukról készült *embeddinglek* egy adatbázisba kerülnek





# FEJLESZTÉSI LEHETŐSÉGEK

- A keresési fázist könnyű automatizálni
- Párhuzamosítani lehet a keresést több közösségi média között
- A demográfiai adatok prediktálásának a pontosítása

# ÖSSZEFOGLALÁS

