



دانشگاه اصفهان  
دانشکده مهندسی کامپیوتر  
گروه هوش مصنوعی

سمینار دفاع از پروژه کارشناسی

موضوع:

طراحی و پیاده‌سازی یک دستیار هوشمند تولید موسیقی بر پایه تحلیل احساسات کاربر

ارائه‌دهنده:

محمد امین کیانی ۴۰۰۳۶۱۳۰۵۲

استاد راهنما: دکتر حسین ماهوش محمدی

استاد داور: دکتر پیمان ادیبی

شهریور ۱۴۰۴



# سرفصل مطالب

✓ مقدمه (مسئله و انگیزه)

✓ مفاهیم

✓ شرح پروژه

✓ نتایج

✓ جمع بندی

✓ پاسخگویی به سوالات





# مقدمه (مسئله و انگیزه)

با پیشرفت فناوری هوش مصنوعی، تحولات چشمگیری در صنعت موسیقی رخ داده است. یکی از این تحولات، فناوری تبدیل احساس به موسیقی است که به کاربر امکان تولید یک آهنگ کامل با ساختار و ملودی مناسب از فضای احساسی را می‌دهد.



❑ نیاز: تولید خودکار موسیقی که با حالت هیجانی کاربر همسو باشد.

❑ کاربردها: بازی/تفریح، موسیقی درمانی، تولید محتوای تطبیقی، فیلم/تبلیغات.

❑ چالش‌ها: تشخیص احساس از چند مدیا است (متن/صورت/صوت)، نگاشت پارامترهای موسیقایی، تولید موسیقی معنادار و قابل قبول انسان.





## ۱- معماری‌های پردازش چندرسانه‌ای:

### CNN\_miniXCEPTION

□ پردازش متن :

معماری BERT (فقط بخش انکودر ترنسفورمر)

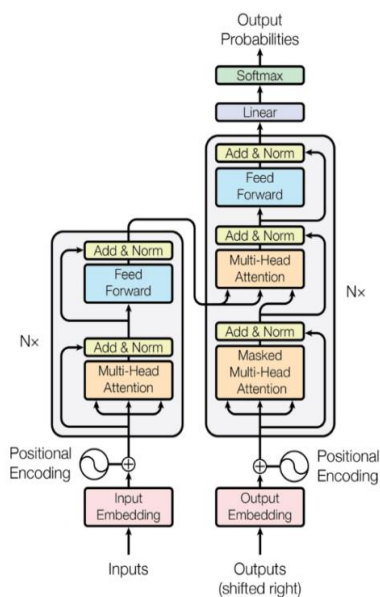
□ مدل پردازش تصویر :

معماری CNN\_miniXCEPTION

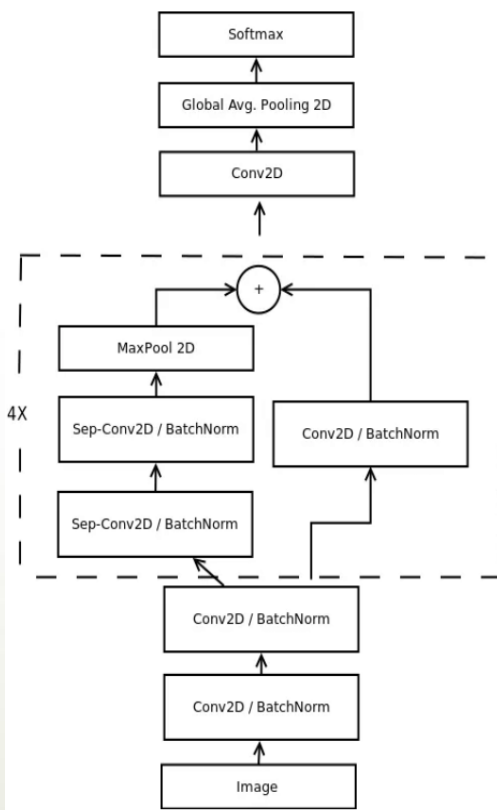
□ مدل پردازش صوت :

معماری Transformer

### Transformer



Encoder Decoder





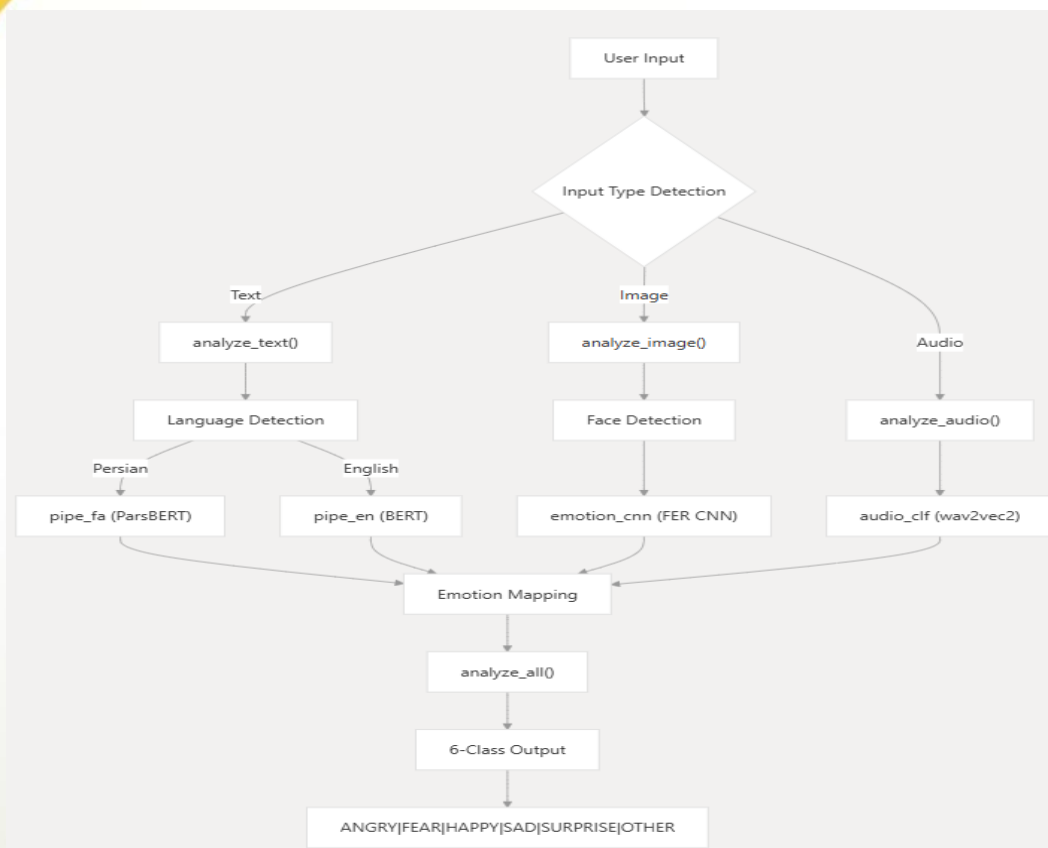
# مفاهیم

## ۲- مدل های پردازش چند رسانه ای (مولتی مودال):

□ مدل های پردازش متن :  
ParsBERT برای فارسی  
BERT برای انگلیسی

□ مدل پردازش تصویر :  
fer2013\_mini\_XCEPTION.1  
02-0.66.hdf5

□ مدل پردازش صوت :  
wav2vec2-base-superb-er







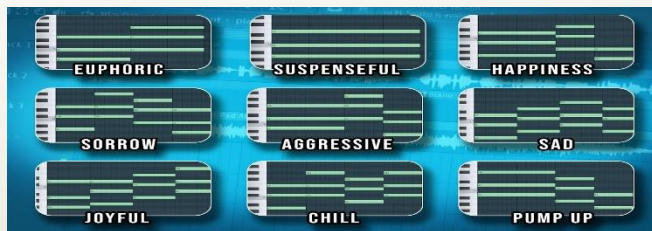
## ۳- تئوری موسیقی:

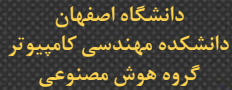
□ آکوردها و آکوردساز (Chord Progression): ترکیبی از چند نت که در یک لحظه نواخته و پایه هارمونی موسیقی را تشکیل می‌دهند. ترتیب دنباله‌ای آکوردها (آکوردساز) ساختار هارمونیک قطعه را شکل می‌دهد.

□ فرم سه‌بخشی A-B-A' : یک ساختار متداول در موسیقی کلاسیک است که شامل سه بخش متقارن می‌شود؛ بخش اول، بخش میانی متفاوت و بخش آخر که معمولاً تکرار یا واریاسیونی از بخش اول است. این فرم شکل متقارن و واضحی به قطعه می‌دهد.

### PROGRESSIONS

Am-C-Am-G	SAD
Em-F-C-Dm	NEUTRAL
G-Am-F-C	EMOTIONAL
C-C-F-G	HAPPY
C-Am-F-G	ROMANTIC
F-C-Am-Em	DARK
C-G-Am-F	EPIC
Dm-E-Dm-C	SPOOKY
C-Ab-Bb	HEROIC





❑ **گام موسیقی (Major/Minor):** گام‌ها مجموعه‌ای از نت‌ها با توالی مشخص‌اند. گام‌های ماژور اغلب بار مثبت (شاد)، و گام‌های مینور بار منفی (غمگین) دارند و نقش مهمی در احساس کلی موسیقی ایفا می‌کنند.

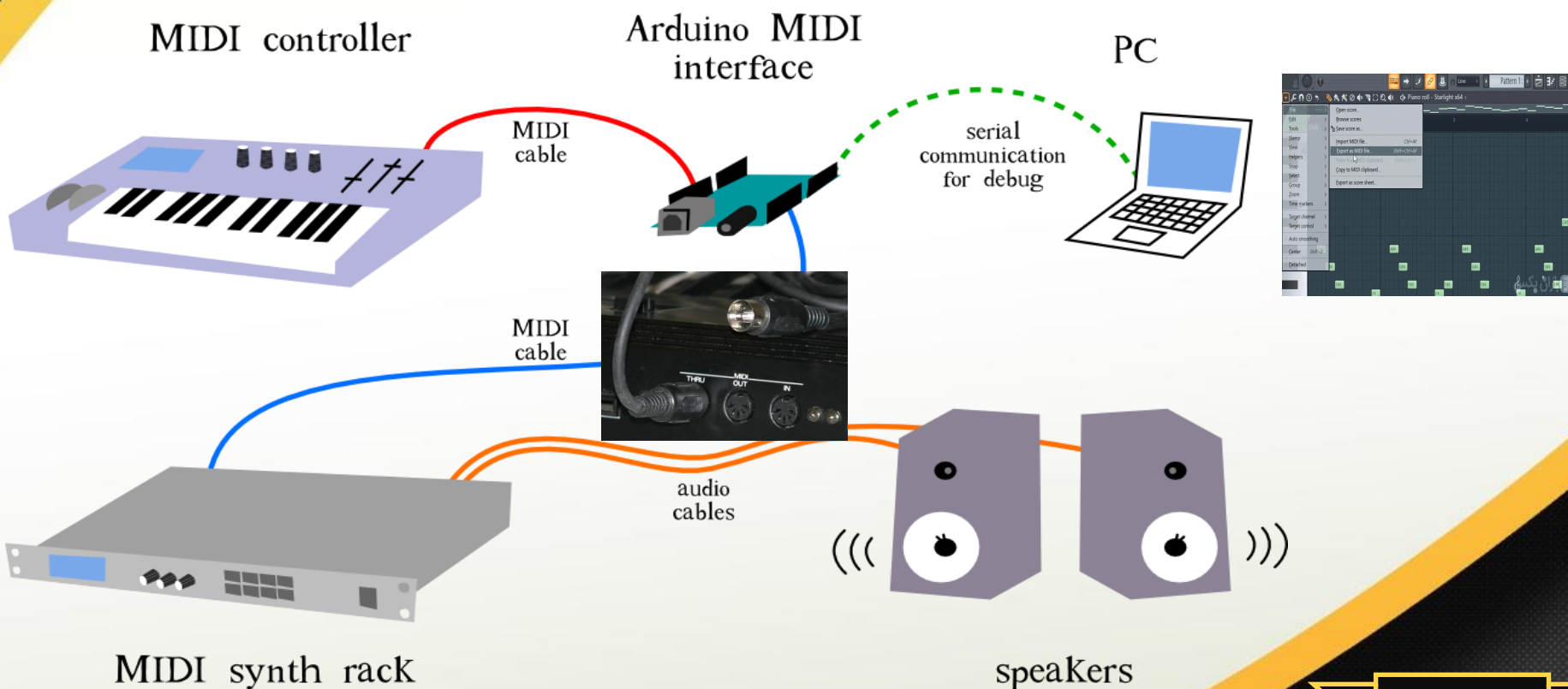
*Key descriptions taken from Charpentier's 'Regles de Composition'; circa 1682*





# مفاهیم

□ رابط دیجیتال آلات موسیقی (**MIDI**): برای انتقال دستورات موسیقی مانند نت‌ها، سرعت و مدت زمان آن‌ها است، نه خود صدا.

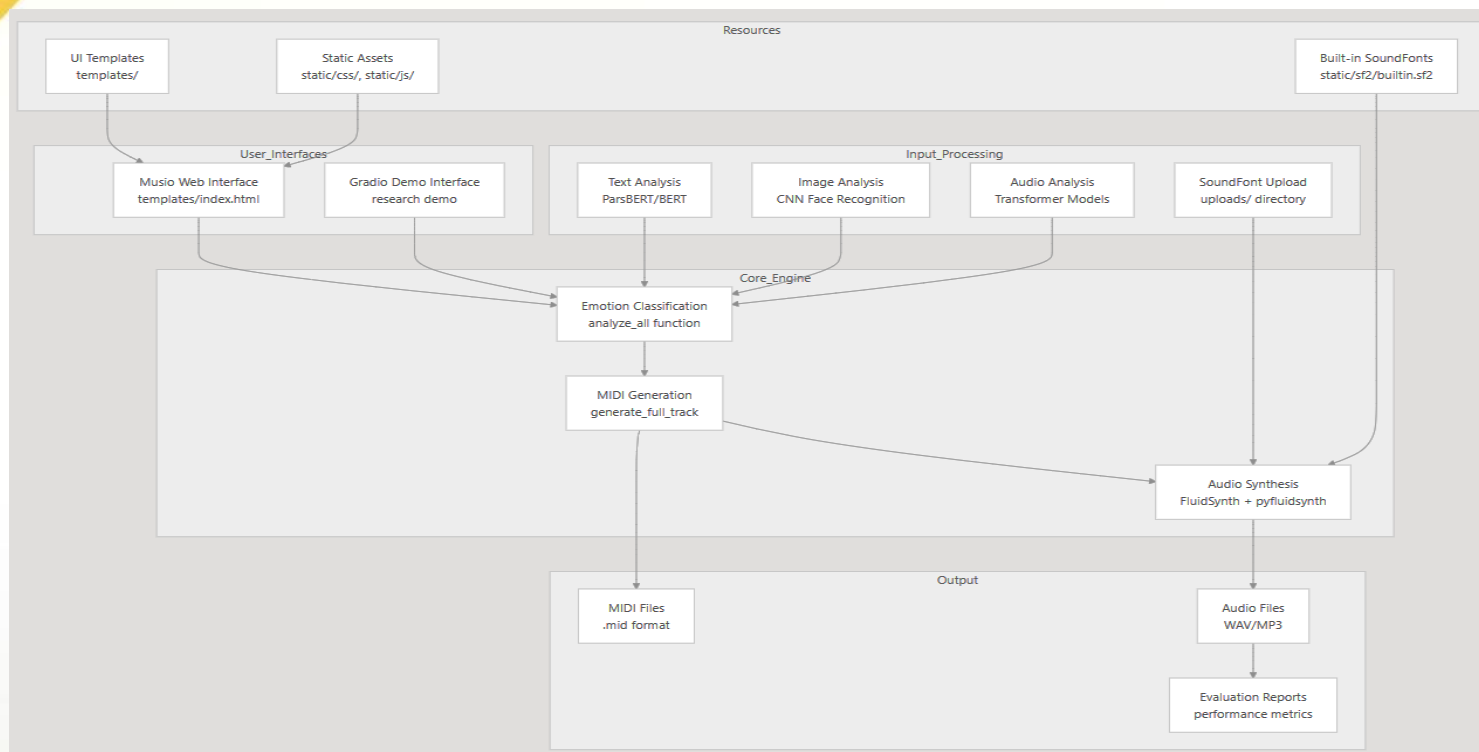






# شرح پروژه

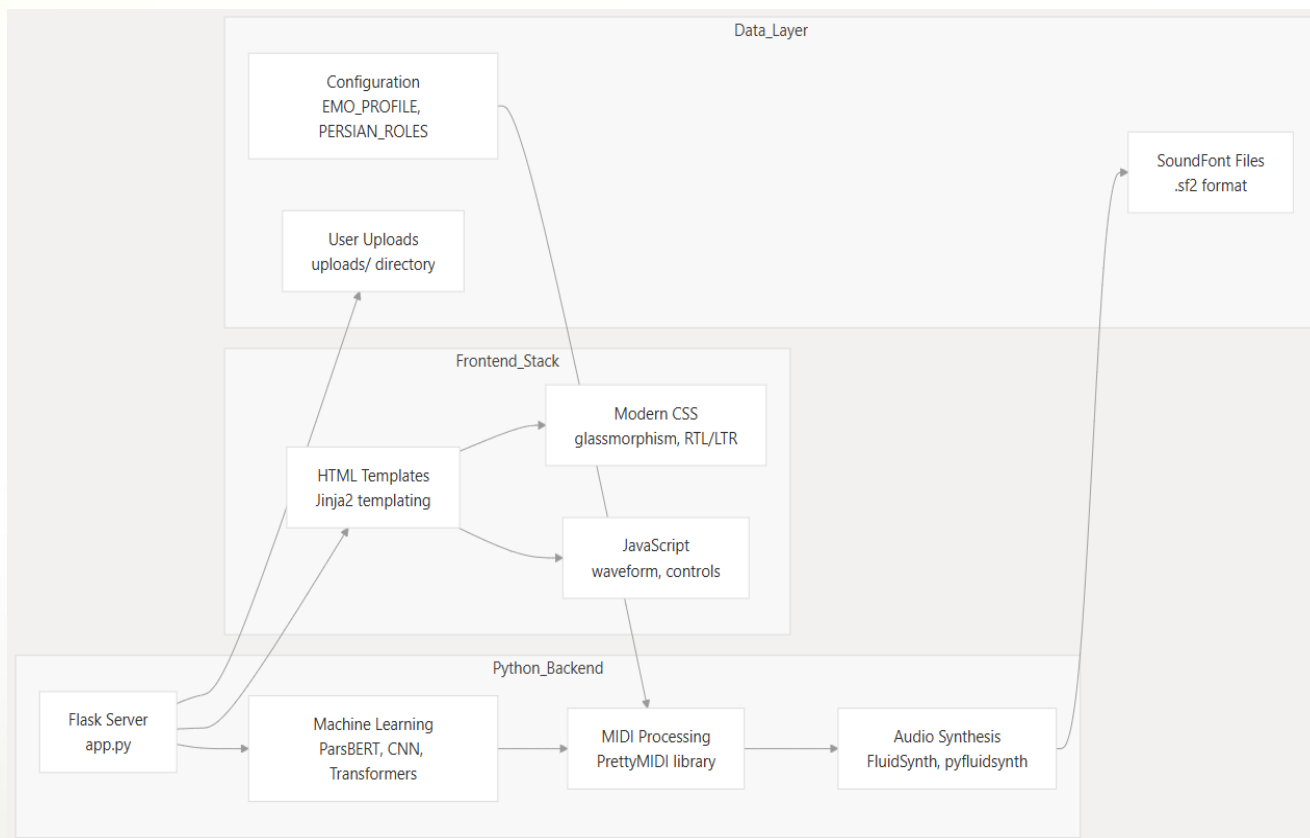
## معماری کلی سیستم





# شرح پروژه

## فناوری‌ها و پیاده‌سازی

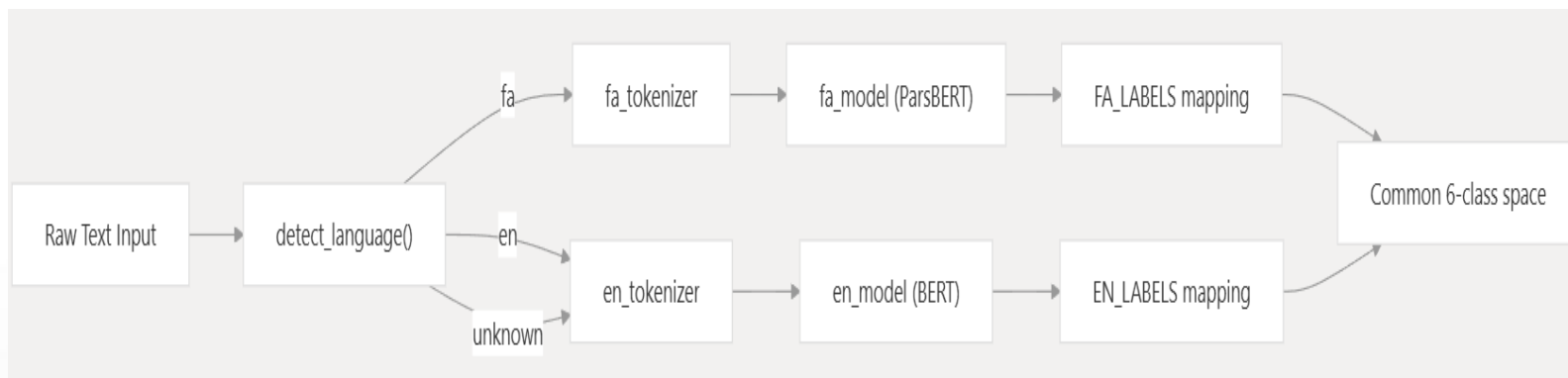




# شرح پروژه

## تحلیل احساس ورودی‌ها

✓ متن (فارسی/انگلیسی)

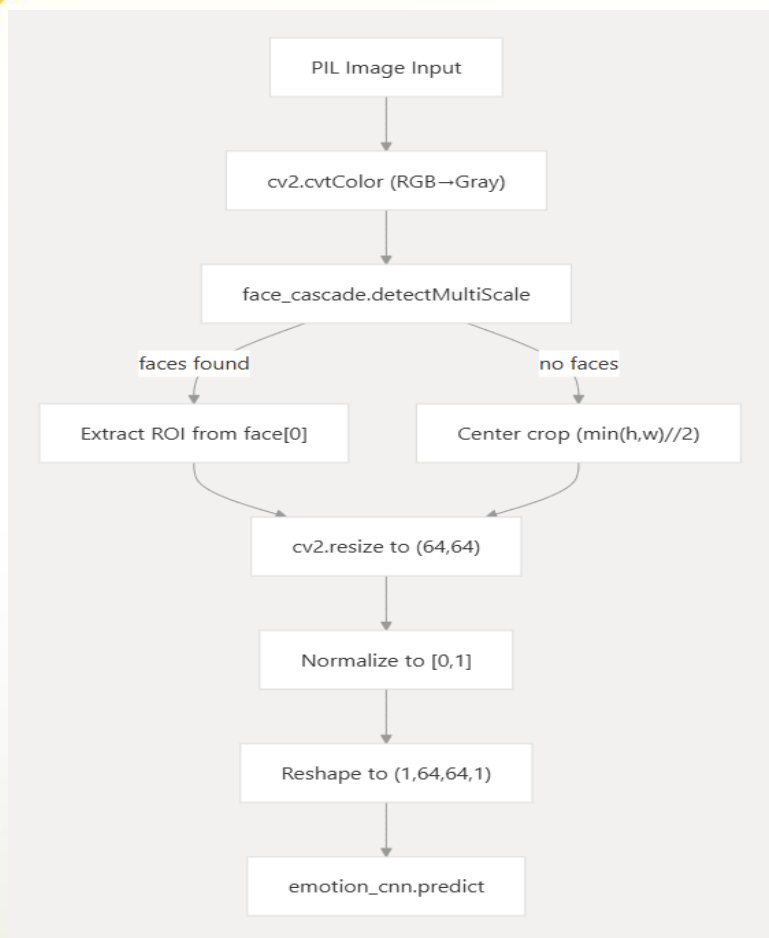




# شرح پروژه

## تحلیل احساس ورودی‌ها

✓ تصویر (چهره)



Parameter	Value
Input Size	64×64 grayscale
Architecture	mini-XCEPTION
Model File	fer2013_mini_XCEPTION.102-0.66.hdf5
Output Classes	['angry', 'disgust', 'fear', 'happy', 'sad', 'surprise', 'neutral']





# شرح پروژه

## تحلیل احساس ورودی‌ها

✓ صوت (ویس)



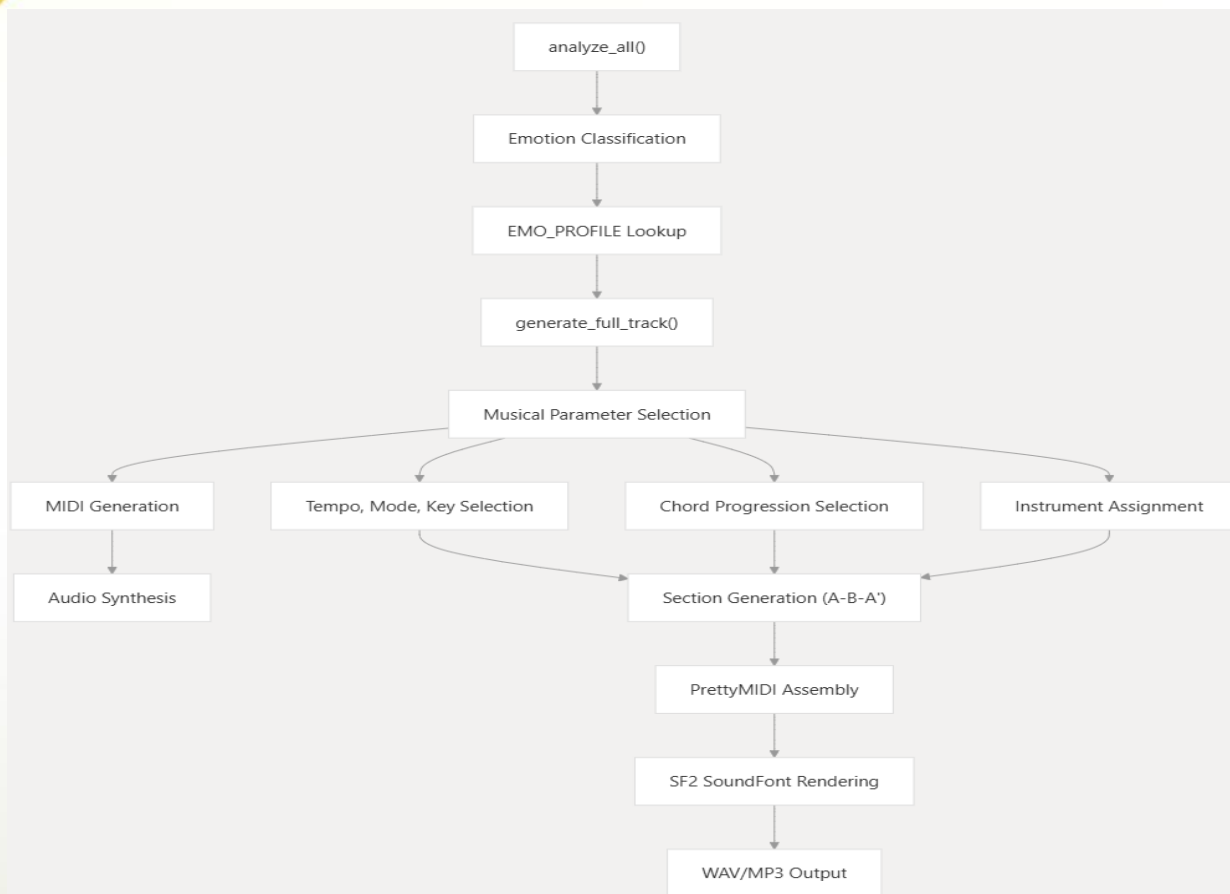
```
audio_pipe = pipeline(  
    "audio-classification",  
    model="superb/wav2vec2-base-superb-er",  
    device=0,  
    top_k=1  
)
```

در ادامه احساس تحلیل شده توسط مدل‌های یادگیری، به کمک سیستم مبتنی بر قانون تئوری آهنگسازی سبب تولید موسیقی نهایی می‌شود.



# شرح پروژه

## جریان آنالیز داده‌ی ورودی

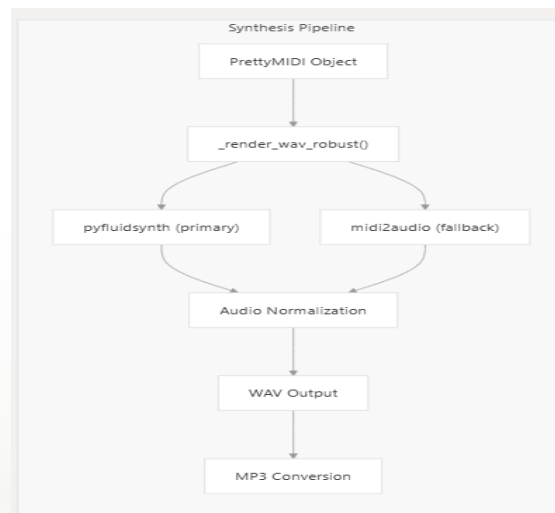
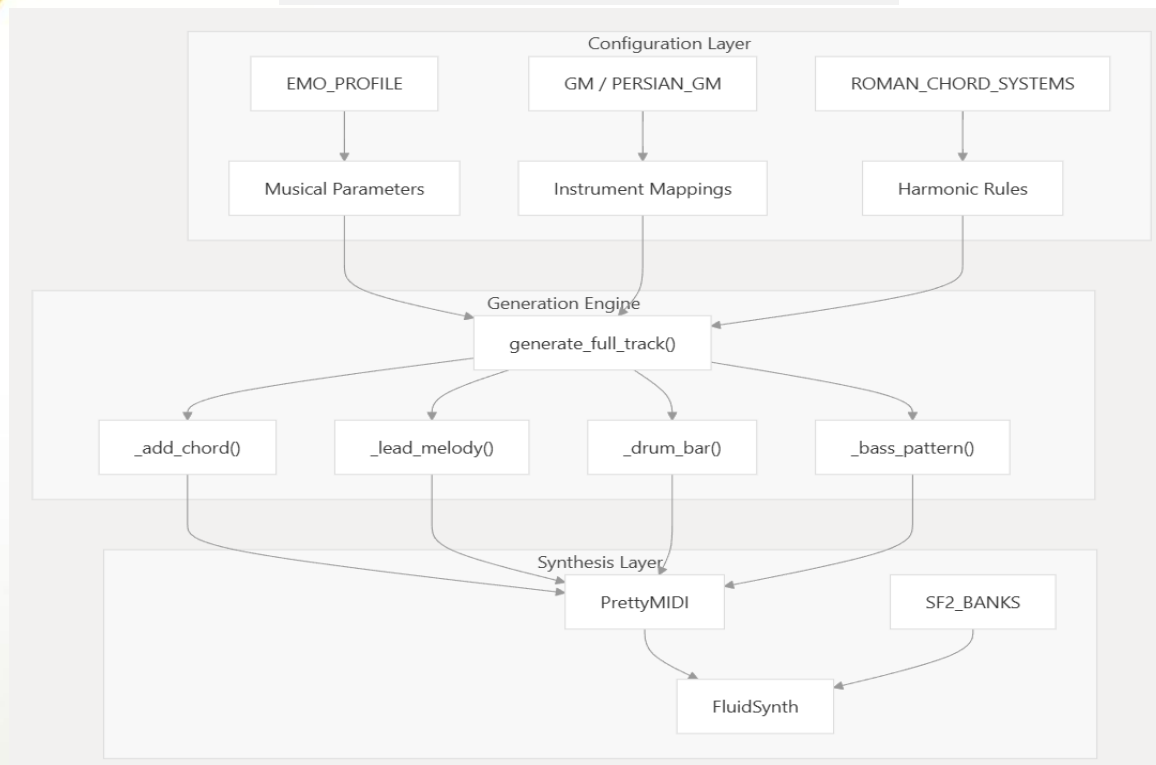




# شرح پروژه

## زیرسیستم‌های تولید موسیقی

### rule-based MIDI music generation engine:



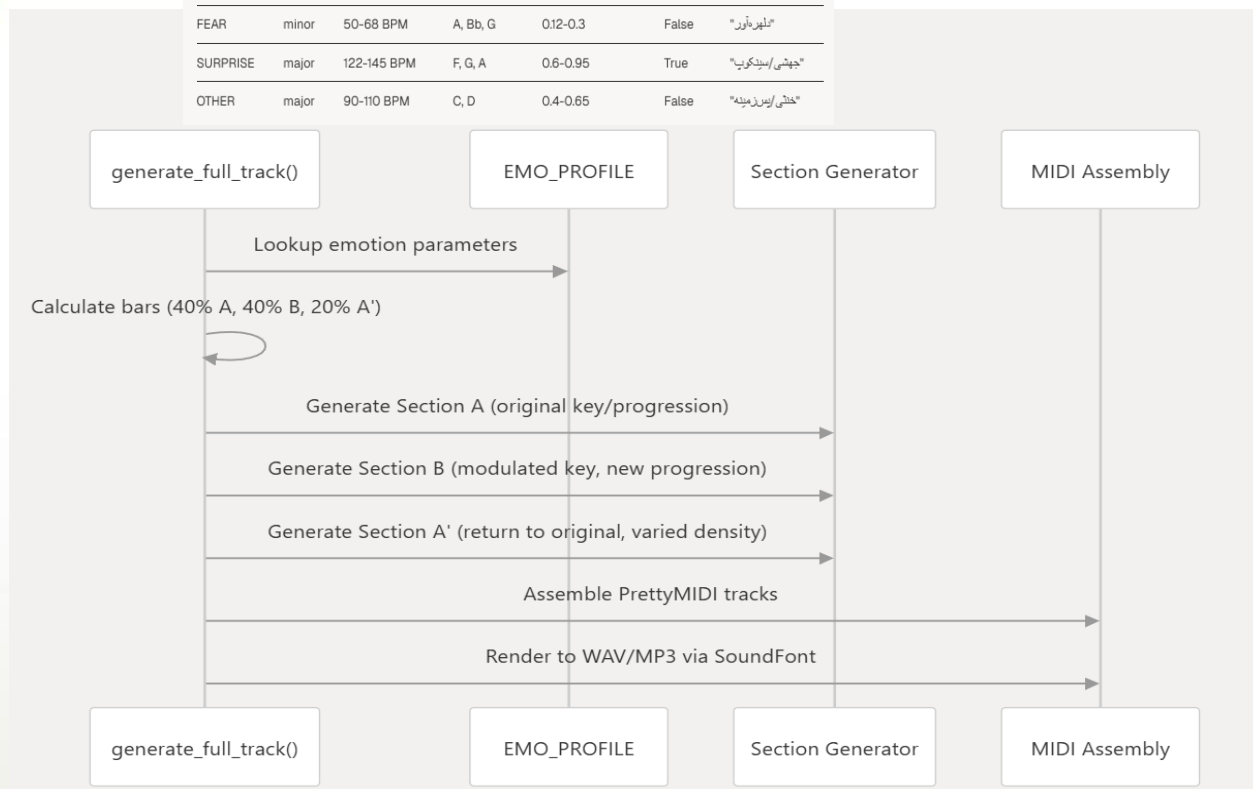


# شرح پروژه

## EMO\_PROFILE:

Emotion	Mode	Tempo Range	Base Keys	Melody Density	Drums	Comment
HAPPY	major	124-142 BPM	C, D, F, G	0.7-0.95	True	"شاد و رهاصان"
SAD	minor	45-60 BPM	A, G, E	0.12-0.28	False	"خیلی آرام و آهسته"
ANGRY	minor	150-175 BPM	D, C, F	0.8-0.96	True	"ترک/بندل نهالچی"
FEAR	minor	50-68 BPM	A, Bb, G	0.12-0.3	False	"تلر داور"
SURPRISE	major	122-145 BPM	F, G, A	0.6-0.95	True	"چپشی/سینکوب"
OTHER	major	90-110 BPM	C, D	0.4-0.65	False	"خشکی/سینکوب"

## ساختار تولید موسیقی

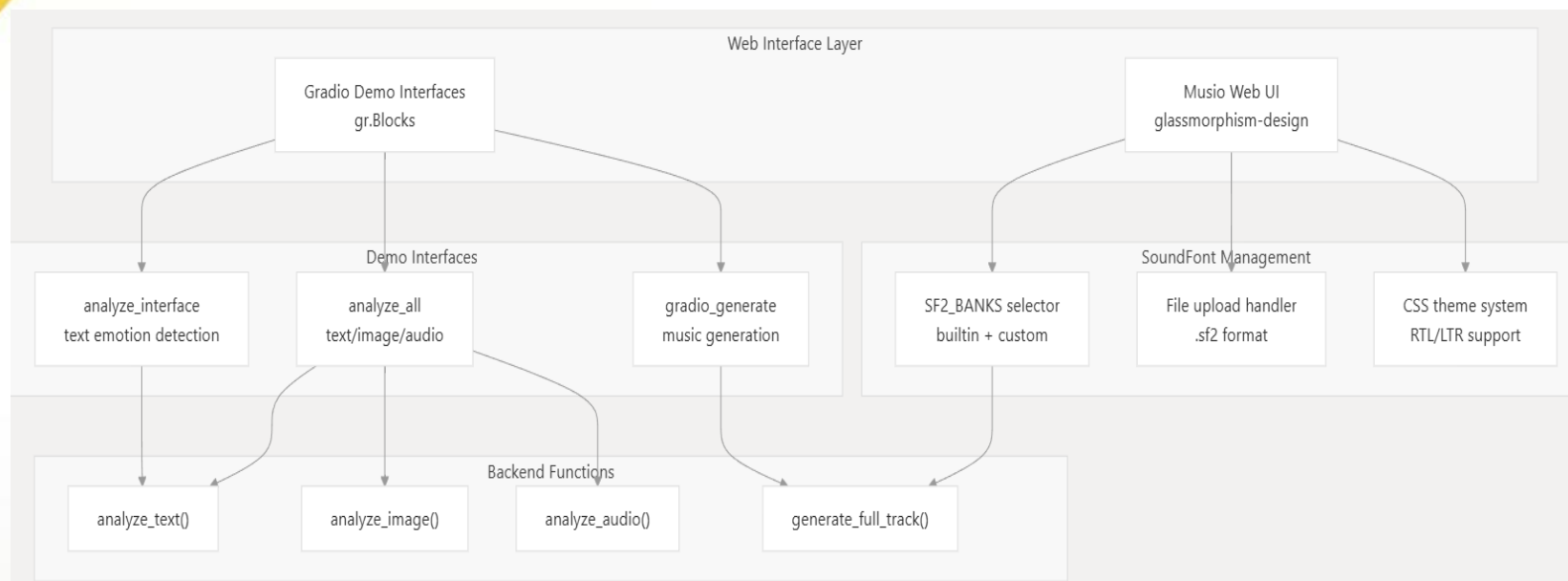






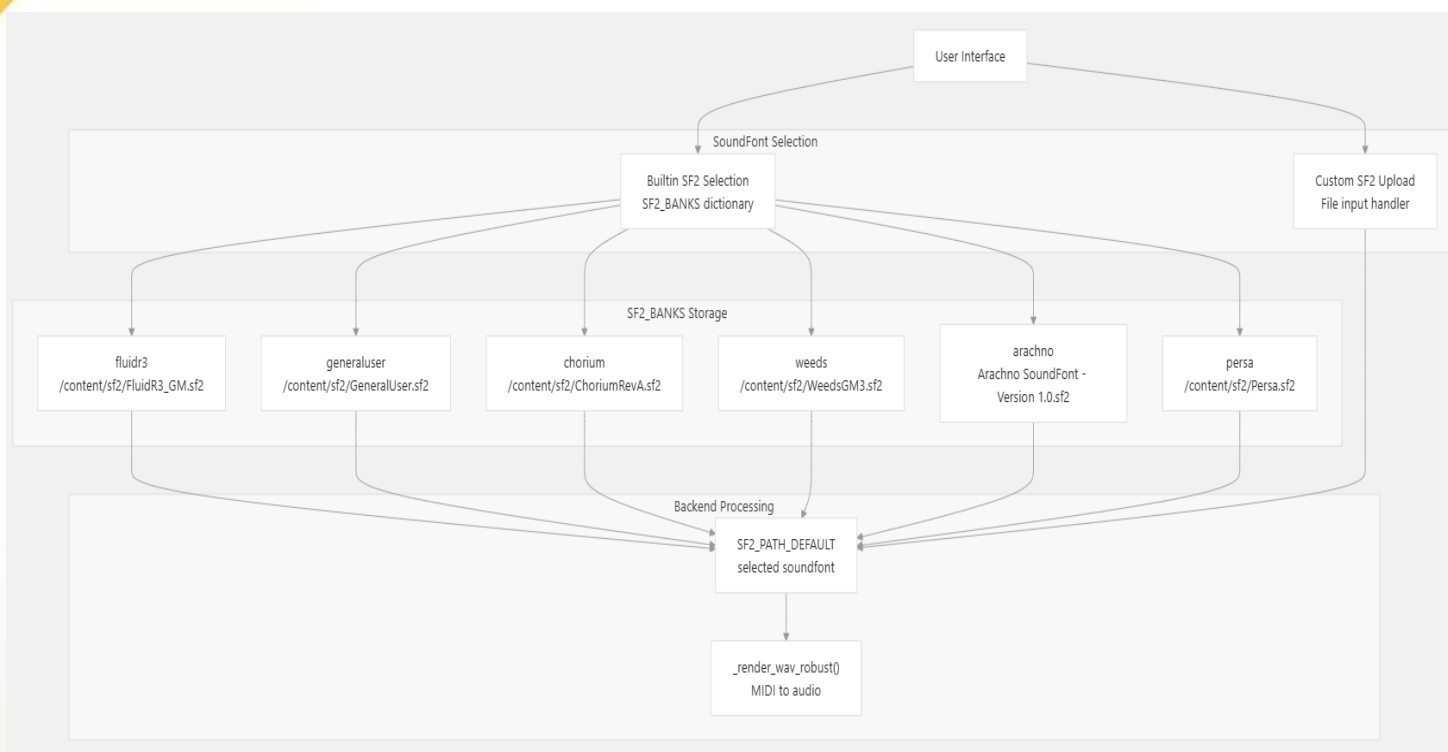
# نتایج

❖ رابط کاربری:





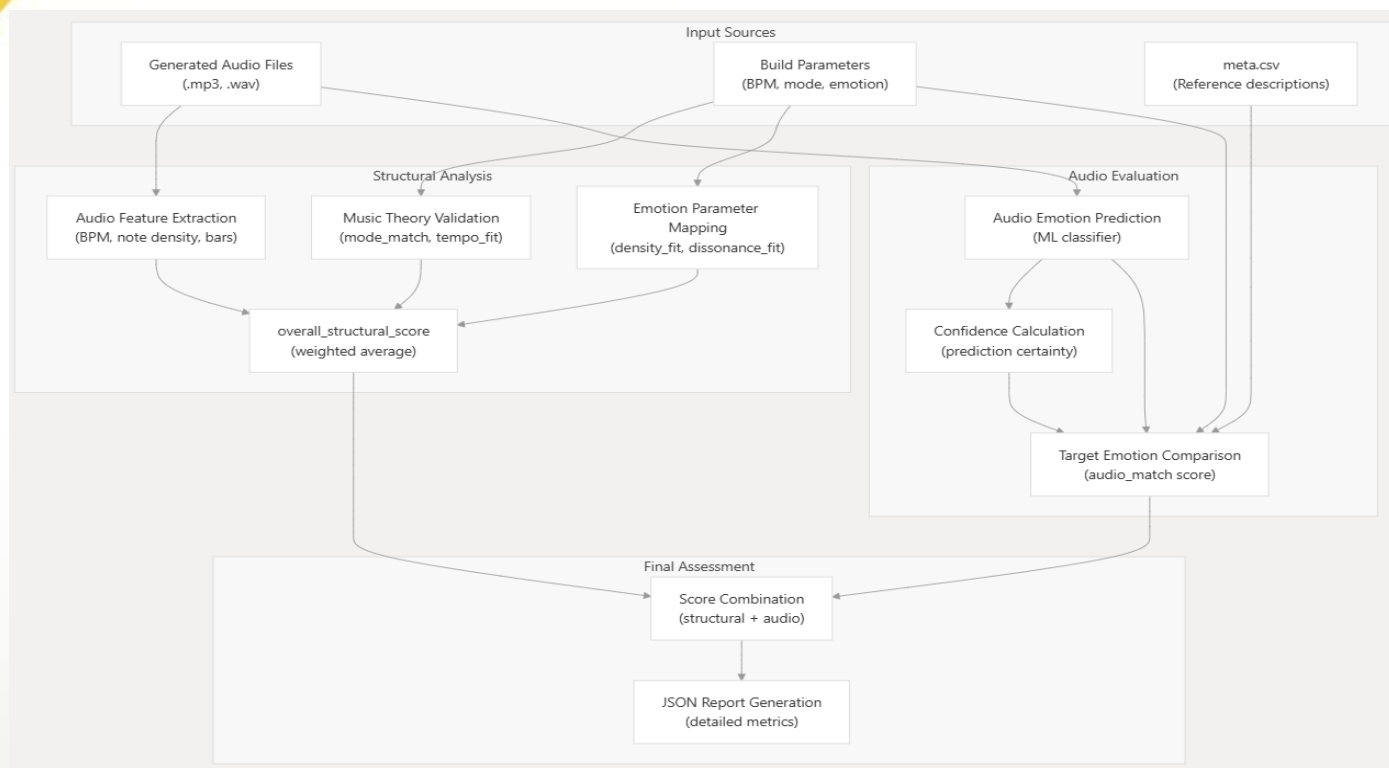
## ❖ فرمت صوتی مبتنی بر نمونه (SF2):





# نتایج

❖ ارزیابی: این سیستم میزان تطابق موسیقی تولید شده با پارامترهای احساسی هدف و محدودیت‌های نظریه موسیقی را کمی می‌کند.





## ❖ نمونه‌ای از ارزیابی برای موسیقی تولیدی:

```
{
  "emotion": "FEAR",
  "style": "global",
  "structural": {
    "bpm": 51,
    "bars": 16,
    "note_density_per_bar": 16.81,
    "mode_match": 1.0,
    "tempo_fit": 1.0,
    "density_fit": 0.0,
    "dissonance_fit": 1.0,
    "syncopation_fit": 0.36,
    "overall_structural_score": 0.686,
    "build_params": {
      "emotion": "FEAR",
      "style": "global",
      "minutes": 1.0,
      "bpm": 51,
      "mode": "minor",
      "key_midi": 58,
      "progression_A": [
        "i",
        "bVI",
        "bVII",
        "i"
      ],
      "progression_B": [
        "i",
        "bVI",
        "bVII",
        "i"
      ],
      "drums": false,
      "comment": "دلهره‌آور",
      "bars_total": 16,
      "sr": 22050
    }
  },
  "audio_eval": {
    "predicted_emotion_audio": "ANGRY",
    "predicted_confidence_audio": 0.969,
    "target_emotion": "FEAR",
    "audio_match": 0.0
  },
  "final_score": 0.515,
  "detected_label": "FEAR",
  "detected_confidence": 0.504,
  "soundfont": "/content/FluidR3_GM.sf2"
}
```

### Audio Generation Evaluation

```
{
  "n_generated": 2,
  "n_reference": 3,
  "metrics": {
    "R@1": 0.5,
    "MeanRank": 1.5,
    "CLAPScore_mean": 0.0675,
    "CLAPScore_median": 0.0675,
    "FAD_VGGish": 352521.70835931465,
    "KAD(CLAP)_mean": 0.0009212652803398669,
    "KAD(CLAP)_std": 0.0022993823513388634,
    "FAD(CLAP)": 1.4007600481266895,
    "Diversity(CLAP)": 0.41101449728012085,
    "MOS_mean": 3.7,
    "MOS_std": 1.1
  }
}
```

### Per-file table

file	tempo_est	key_index_est	text	CLAP_text_audio
/content/sad_persian_1.0min_1755642423.wav	117.19	11	very slow, mournful, crying, minimal, sparse	0.062850
/content/sad_persian_1.0min_1755642483.wav	96.98	0	very slow, mournful, crying, minimal, sparse	0.072201

### MOS (fake) table

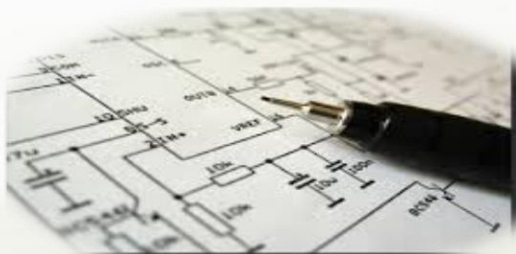
file	mos_1	mos_2	mos_3	mos_4	mos_5	mos_6	mos_7	mos_8	mos_9	mos_10	mos_11	mos_12	mos_13	mos_14	mos_15
/content/sad_persian_1.0min_1755642423.wav	2	5	4	3	3	5	2	4	2	2	4	5	4	5	4
/content/sad_persian_1.0min_1755642483.wav	5	4	2	5	3	4	3	2	5	5	4	3	5	4	3





# جمع بندی

- در این پروژه با ترکیب مدل‌های یادگیری و روش‌های مبتنی بر قوانین تولید موسیقی، توانستیم ملودی‌های متناسب با احساسات چندرسانه‌ای تولید کنیم.
- نتایج نشان داد که با رعایت قواعد هارمونی و فرم موسیقی (A-B-A')، ارتباط معنی‌دار بین احساس تحلیل‌شده و موسیقی برقرار می‌شود. نوآوری اصلی این پروژه در یکپارچه‌سازی تحلیل احساسی فارسی با تولید خودکار موسیقی و استفاده از فرم‌های مختلف موسیقی بود.
- **پیشنهادهای:** گسترش مجموعه داده‌های آموزشی (ورودی و موسیقی)، بهبود الگوریتم تولید (مثلاً استفاده از مدل‌های پیشرفته‌تر یا یادگیری تقویتی)، و انجام آزمایش‌های شنیداری بیشتر با شرکت‌کنندگان بومی.





# پرسش و پاسخ

با سپاس از توجه شما





دانشگاه اصفهان  
دانشکده مهندسی کامپیوتر  
گروه هوش مصنوعی

# پرسش و پاسخ

## سوال



Aminkiani82@gmail.com



<https://github.com/M-Amin-Kiani/bachelor-s-dissertation/>