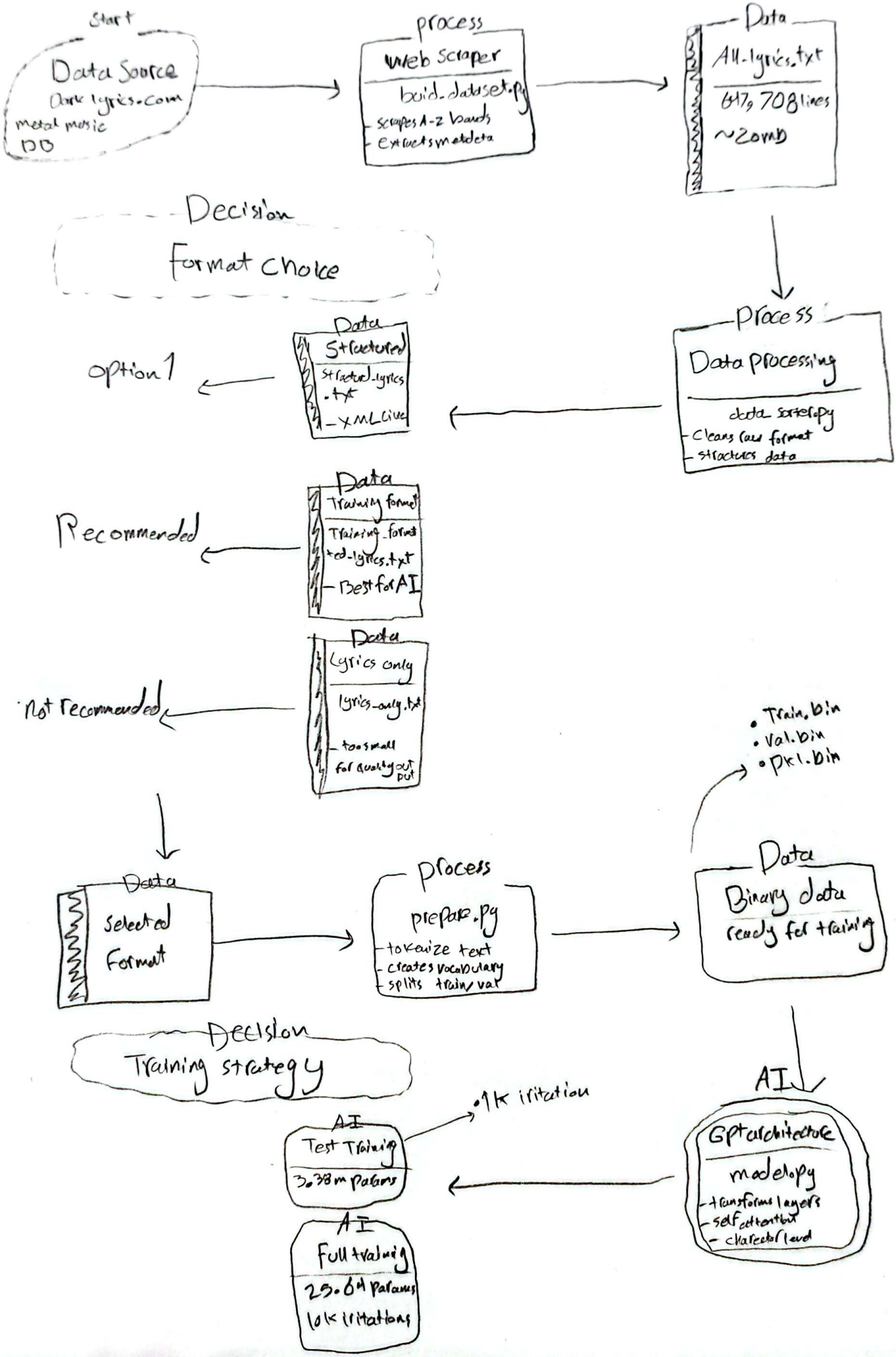
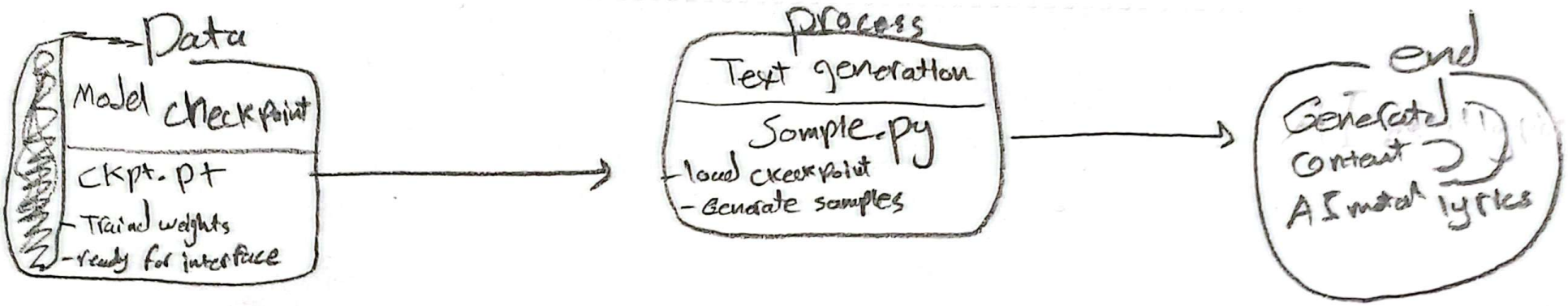

Start
Data Source
Dark lyrics.com
metal music
DB
process
Web scraper
build_dataset.py
- scrapes A-z bands
- extracts metadata
Data
All_lyrics.txt
617, 708 lines
~20mB
Decision
Format choice
Process
Data processing
data_sorter.py
- Cleans raw format
- structures data
Data
Structured
structured_lyrics.txt
- XML Like
option 1
Data
Training format
Training_format_red_lyrics.txt
- Best for AI
Recommended
Data
Lyrics only
lyrics_only.txt
- too small for quality output
not recommended
Data
Selected Format
process
prepare.py
- tokenize text
- creates vocabulary
- splits train/val
• Train.bin
• val.bin
• pkl.bin
Data
Binary data ready for training
Decision
Training strategy
AI
Test Training
3.38m params
•1k iteration
AI
Full training
25.6M params
10k iterations
AI
GPT architecture
model.py
- transforms layers
- self_attention
- character level


Data
Model checkpoint
ckpt.pt
- Trained weights
- ready for interface
process
Text generation
Sample.py
- load checkpoint
- Generate samples
end
Generated
content
AI metal
lyrics