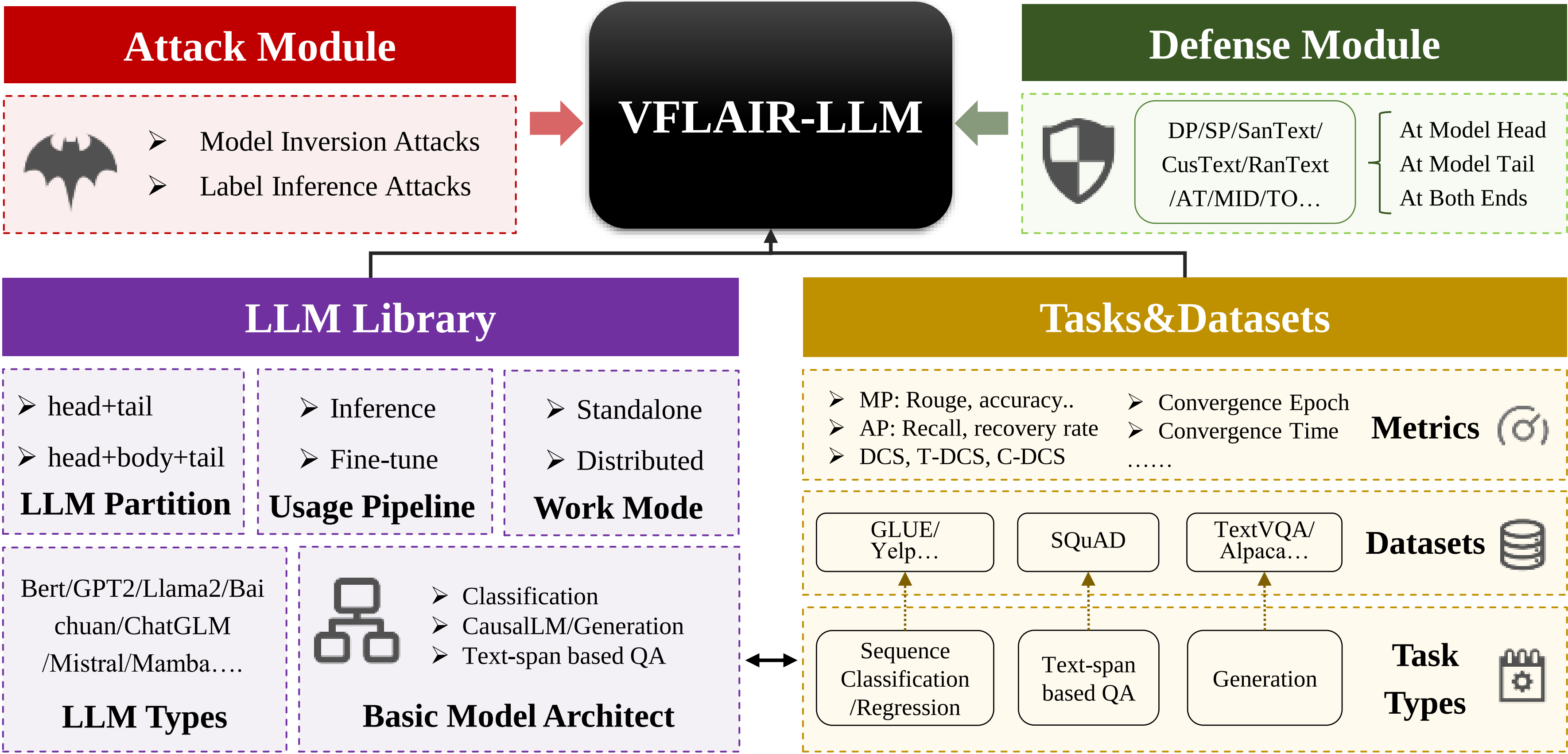

Attack Module

- Model Inversion Attacks
- Label Inference Attacks

VFLAIR-LLM

Defense Module

DP/SP/SanText/
CusText/RanText
/AT/MID/TO…

At Model Head
At Model Tail
At Both Ends

LLM Library

- head+tail
- head+body+tail

LLM Partition

- Inference
- Fine-tune

Usage Pipeline

- Standalone
- Distributed

Work Mode

Bert/GPT2/Llama2/Baichuan/ChatGLM
/Mistral/Mamba….

LLM Types

- Classification
- CausalLM/Generation
- Text-span based QA

Basic Model Architect

Tasks&Datasets

- MP: Rouge, accuracy..
- AP: Recall, recovery rate
- DCS, T-DCS, C-DCS
- Convergence Epoch
- Convergence Time

……

Metrics

GLUE/
Yelp…

SQuAD

TextVQA/
Alpaca…

Datasets

Sequence
Classification
/Regression

Text-span
based QA

Generation

Task
Types