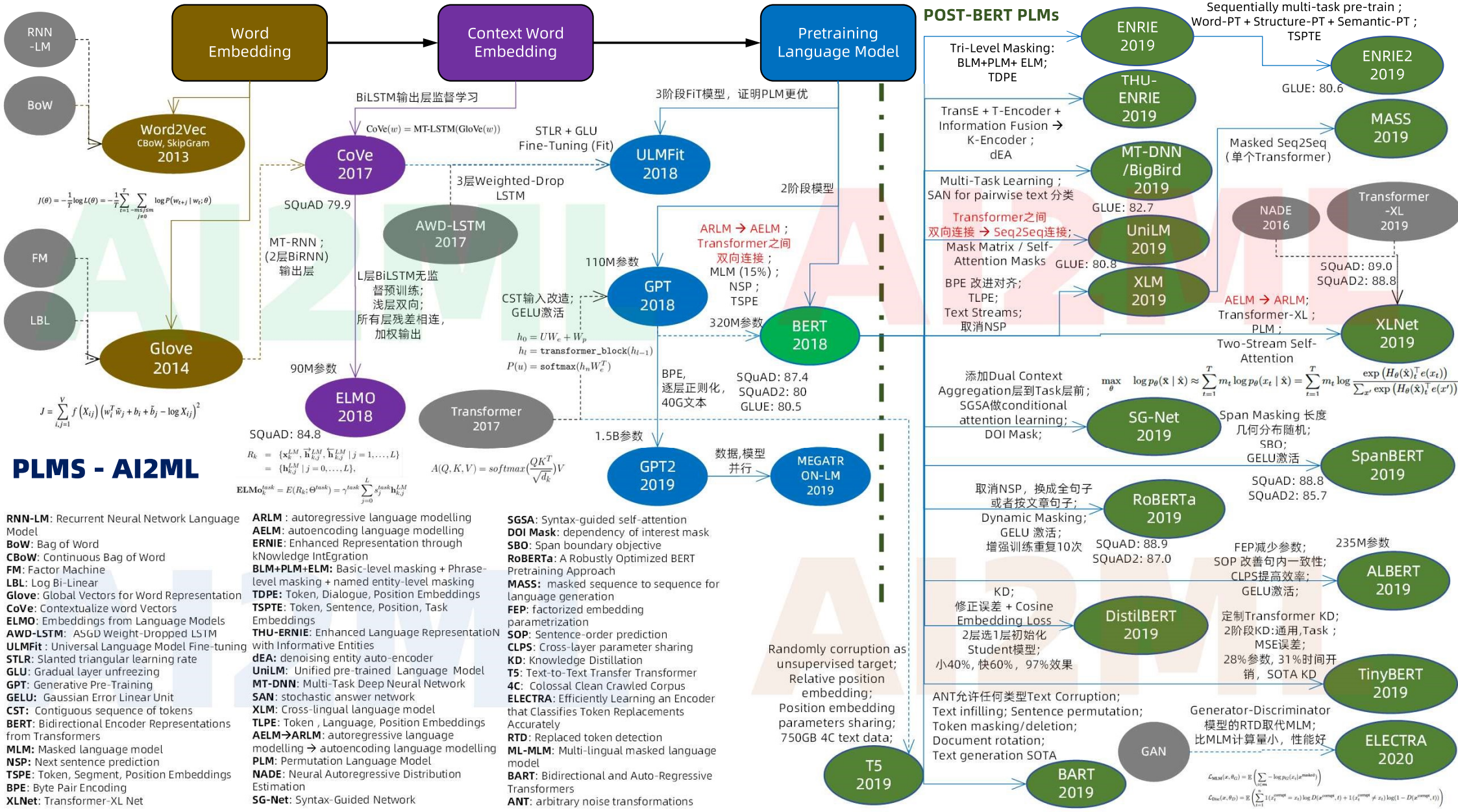

PLMS - AI2ML

RNN-LM: Recurrent Neural Network Language Model
BoW: Bag of Word
CBoW: Continuous Bag of Word
FM: Factor Machine
LBL: Log Bi-Linear
Glove: Global Vectors for Word Representation
CoVe: Contextualize word Vectors
ELMO: Embeddings from Language Models
AWD-LSTM: ASGD Weight-Dropped LSTM
ULMFit : Universal Language Model Fine-tuning
STLR: Slanted triangular learning rate
GLU: Gradual layer unfreezing
GPT: Generative Pre-Training
GELU: Gaussian Error Linear Unit
CST: Contiguous sequence of tokens
BERT: Bidirectional Encoder Representations from Transformers
MLM: Masked language model
NSP: Next sentence prediction
TSPE: Token, Segment, Position Embeddings
BPE: Byte Pair Encoding
XLNet: Transformer-XL Net

ARLM : autoregressive language modelling
AELM: autoencoding language modelling
ERNIE: Enhanced Representation through kNowledge IntEgration
BLM+PLM+ELM: Basic-level masking + Phrase-level masking + named entity-level masking
TDPE: Token, Dialogue, Position Embeddings
TSPTE: Token, Sentence, Position, Task Embeddings
THU-ERNIE: Enhanced Language RepresentationN with Informative Entities
dEA: denoising entity auto-encoder
UniLM: Unified pre-trained Language Model
MT-DNN: Multi-Task Deep Neural Network
SAN: stochastic answer network
XLM: Cross-lingual language model
TLPE: Token , Language, Position Embeddings
AELM→ARLM: autoregressive language modelling → autoencoding language modelling
PLM: Permutation Language Model
NADE: Neural Autoregressive Distribution Estimation
SG-Net: Syntax-Guided Network

SGSA: Syntax-guided self-attention
DOI Mask: dependency of interest mask
SBO: Span boundary objective
RoBERTa: A Robustly Optimized BERT Pretraining Approach
MASS: masked sequence to sequence for language generation
FEP: factorized embedding parametrization
SOP: Sentence-order prediction
CLPS: Cross-layer parameter sharing
KD: Knowledge Distillation
T5: Text-to-Text Transfer Transformer
4C: Colossal Clean Crawled Corpus
ELECTRA: Efficiently Learning an Encoder that Classifies Token Replacements Accurately
RTD: Replaced token detection
ML-MLM: Multi-lingual masked language model
BART: Bidirectional and Auto-Regressive Transformers
ANT: arbitrary noise transformations