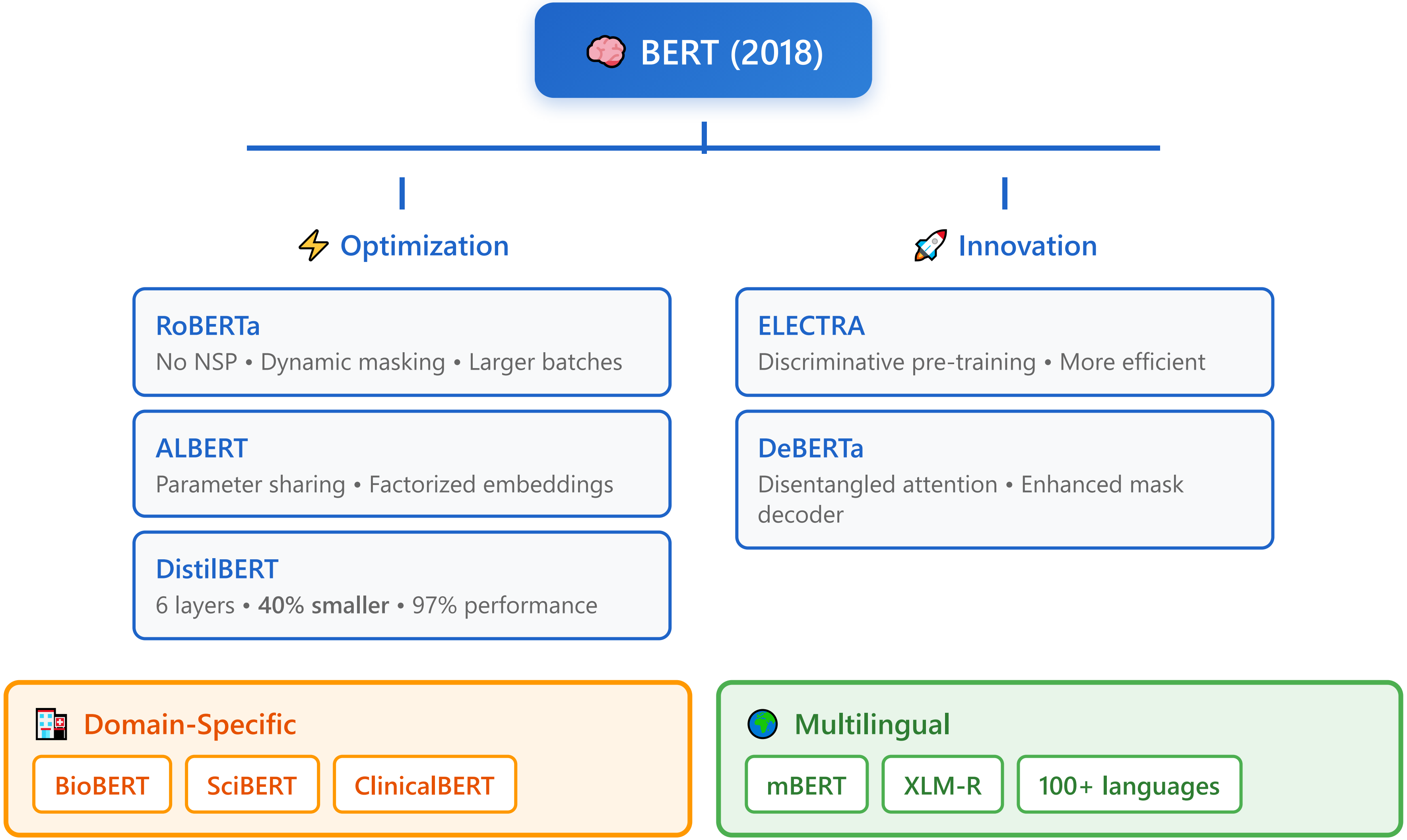
BERT Family: Variants and Improvements



🧠 BERT (2018)

⚡ Optimization

RoBERTa
No NSP • Dynamic masking • Larger batches

ALBERT
Parameter sharing • Factorized embeddings

DistilBERT
6 layers • **40% smaller** • 97% performance

🚀 Innovation

ELECTRA
Discriminative pre-training • More efficient

DeBERTa
Disentangled attention • Enhanced mask decoder

🏥 Domain-Specific

BioBERT | SciBERT | ClinicalBERT

🌍 Multilingual

mBERT | XLM-R | 100+ languages