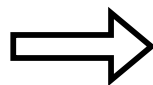
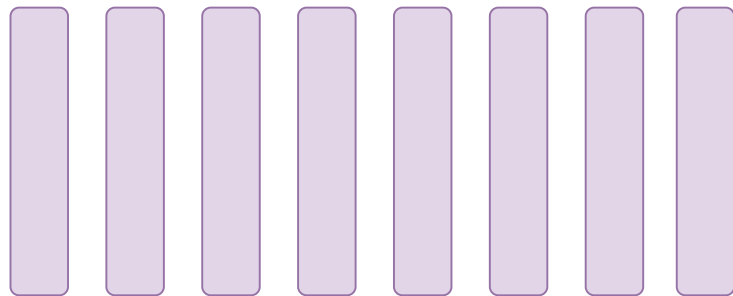


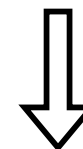
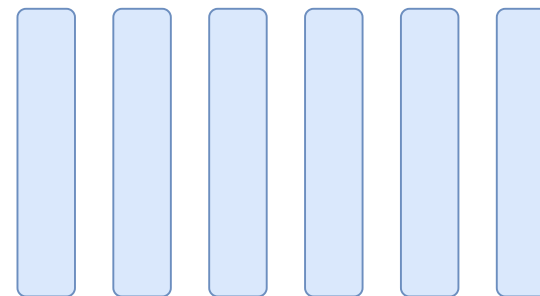
Hidden representations




Vec2Vec
Transformation



New Hidden representations



Decoder of 
Detoxification System for x



Output in X

Детоксифицированный
текст на англ

Encoder of
LLM for Y

Input in Y

Какой-то токсичный
текст на русс

