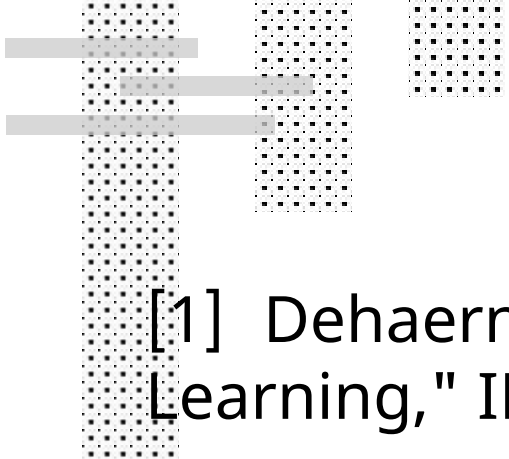# Machine Learning for Code Generation: Literature Review
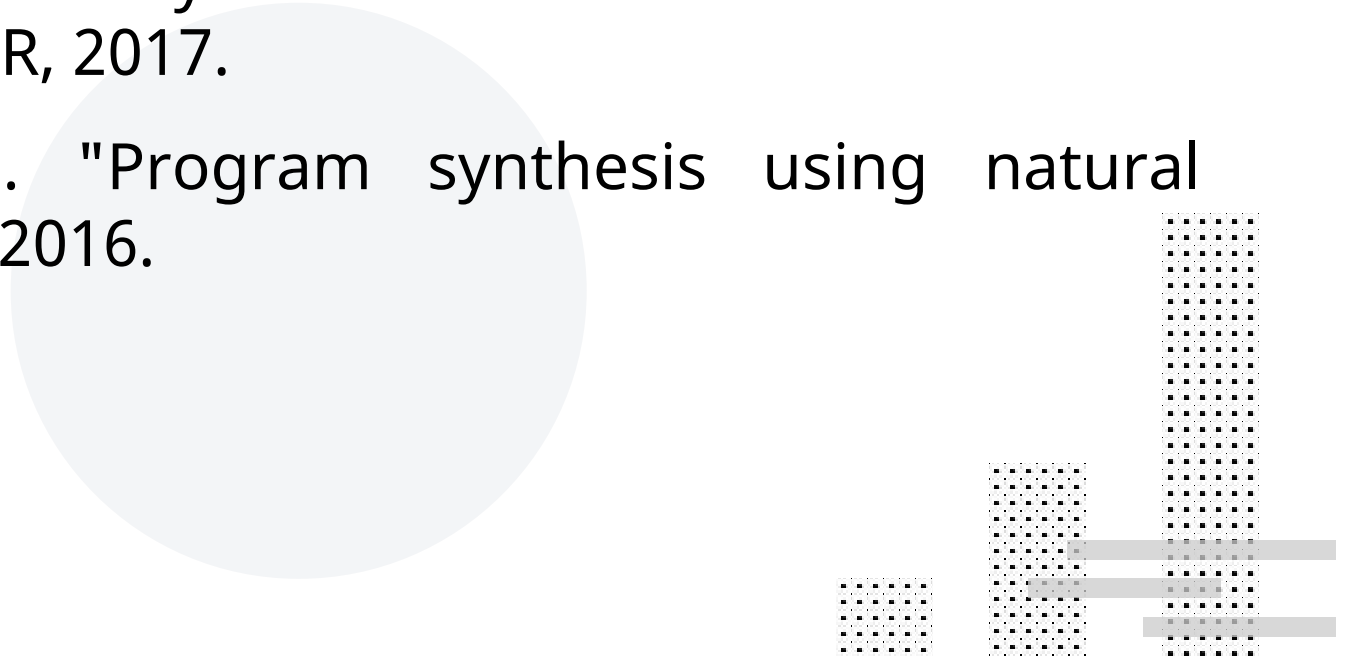## Nikita Safronov

12/4/2024

[1] Dehaerne et al. "Code Generation Using Machine Learning," IEEE Access, 2022.

[2] Yin, Neubig. "A Syntactic Neural Model for Code Generation," CoRR, 2017.

[3] Desai et al. "Program synthesis using natural language," ACM, 2016.

Which ML models are used?

- RNN — Sequential data.
- CNN — Local patterns.
- Transformer — Long-range dependencies.

Where does training data come from?

- 62% — GitHub (large datasets).
- 20% — StackOverflow (NL-code pairs).
- 18% — Other sources (custom datasets).

Thanks for the Attention!