

Data Principles

- Ensure the original data is correct.
- Fix the root data instead of generated data.
- Use a lot of checking script to ensure data correctness.
- Check the correctness and consistency of data.
- Some data needs to be completely accurate, while other data can be approximately correct. For example, translations are acceptable even if they contain occasional typos.
- Use quantitative metrics to measure data quality.
- Wrong data will lead to more wrong data.
- Data includes numbers, strings, text, data structures, images, videos, code, files, and other information types.
- Keep original data in experiments for physics, science, and computer science.
- Data should be reproducible and repeatable.