Lecture 15:

Simulating Virtual Worlds (Two paper discussions)

Visual Computing Systems
Stanford CS348K, Spring 2023

Most of today's class was in-class discussion
(No slides)

Today

- Discussion of the high performance many world simulation paper (“Madrona paper”)
- Discussion of the generative agents paper
- In both cases, a focus on the principles of evaluation

What is the purpose of evaluation?

- Any scientific paper proposes a hypothesis. You can think of a hypothesis as a claim that could be falsified with the right data

- The goal of the evaluation is to:
  - Show that the hypothesis is not falsified by experiments
  - Show that the hypothesis is falsified in certain situations, providing bounds on the hypothesis
  - Provide the reader a better understanding of the trends, limits, value of the hypothesis

From our lecture 2 reading…

The scientific hypothesis is that the proposed organization will lead to better X.

Where might be one or more of performance, developer productivity, maintainability, reliability, safety, etc.

The key insight: many good systems papers put forth a novel organizing observation about the structure of a problem.

Given stated goals and constraints, a systems paper will often propose a formulation of a problem that facilitates meeting these requirements. In other words, many systems papers make the argument:

It is useful to think about the problem in terms of these structures (e.g., using these abstractions, these representations, or by factoring the problem into these modules), because when one does, there are compelling benefits.

Benefits might take the form of: improved system performance (or scaling), enhanced programmer productivity, greater system extensibility/generality, or the ability to provide new application-level capabilities that have not been possible before.

Identifying useful problem structure often forms the central intellectual idea of a systems paper. As in other areas of computer graphics, elegant conceptual insights can be summarized in a few sentence

From our lecture 2 reading…

Good systems papers highlight key design decisions (and discuss alternatives to those decisions).

Given a set of requirements, a systems architect is usually faced with a variety of solution strategies. For example, a performance requirement could be solved through algorithmic innovation, through the design of new specialized hardware, or both (modifying an existing algorithm to better map to existing parallel hardware). Alternatively, the path to better performance might best go through narrowing system scope to a smaller domain of tasks. A productivity goal might be achieved through the design of new programming abstractions, which might be realized as a new domain specific language, or via a library implemented in an existing system.

As a result, a systems paper author must identify the key choices made in architecting their system, and elaborate on the rationale for these design decisions. Doing so typically involves discussing the palette of potential alternatives, and providing an argument for why the choices made are a preferred way to meet the stated requirements and design goals. It is not sufficient to merely describe the path that was taken without saying why was deemed a good one.

From our lecture 2 reading…

The evaluation: were the key design decisions responsible for meeting the stated requirements and goals?

If a paper clearly describes a system's goals and constraints, as well as articulates key system design decisions, then the strategy for evaluating the system is to provide evidence that the described decisions were responsible for meeting the stated goals.

Particularly when a system's evaluation focuses on performance, it is tempting to compare the proposed system's end-to-end performance against that of competing alternative systems. While such an evaluation demonstrates that performance goals were met, it is equally (and sometimes more) important to conduct experiments that specifically assess the benefit of key optimizations and design decisions. Evaluation of why success was achieved is necessary to verify that the central claims of the paper are sound. Failing to perform this evaluation leaves open the possibility that the success of the system is due to other factors. (e.g., high-quality software engineering), than the proposed key ideas.