AI in Software Development

CMSC398W: Practical Tools For Efficient Development

Mohammad Durrani

June 22, 2026

What We're Covering Today

Four Topics

Topic	Focus
Productivity	When AI helps and when it does not
Specification	Writing clear instructions so the AI does what you mean
Context	Giving the AI the right information
Code Review	Checking that the output is actually correct

Access is Easy, Shipping Actual Code is Hard

The Tools Are Here

Installing Claude Code, Gemini CLI, Cursor, or Copilot takes five minutes and generating code is trivial.

Access is Easy, Shipping Actual Code is Hard

The Tools Are Here

Installing Claude Code, Gemini CLI, Cursor, or Copilot takes five minutes and generating code is trivial.

So Why Study This?

If generating code is easy, why dedicate a lecture to it? Because generating code is not the same as shipping working software. Without proper technique, AI tools can actually slow you down.

AI Making Undesired Changes

The Setup

You ask the AI to fix a small CSS bug on a button. It fixes the bug, but also "cleans up" the rest of the component to use a slightly different state management pattern.

AI Making Undesired Changes

The Setup

You ask the AI to fix a small CSS bug on a button. It fixes the bug, but also "cleans up" the rest of the component to use a slightly different state management pattern.

The Catch

Did you just fix a bug, or did you just adopt code you no longer understand?
The button works, but the AI silently removed the edge-case handling for offline users. You have lost cognitive control over the file. If you spend an hour untangling the AI's "helpful" refactor to restore the old logic, you lost time.

Is AI Actually Faster?

Look At The Numbers

Studies show mixed results. A 2024 Copilot study showed junior developers gaining 27-39% productivity, while seniors gained 8-13%.[1]

But a 2024 Uplevel study found AI assistants actually increased cycle time and introduced more bugs in some enterprise teams.

[1] Demirer et al, The Effects of Generative AI on High Skilled Work, 2024.

Is AI Actually Faster?

Look At The Numbers

Studies show mixed results. A 2024 Copilot study showed junior developers gaining 27-39% productivity, while seniors gained 8-13%.[1]
But a 2024 Uplevel study found AI assistants actually increased cycle time and introduced more bugs in some enterprise teams.

[1] Demirer et al, The Effects of Generative AI on High Skilled Work, 2024.

How Do We Estimate Time Saved?

Net time saved equals AI generation time minus the time you spend fixing its mistakes.

The Techniques Stay, The Models Change

Model Evolution

Models get better every month, meaning a tool that fails at a task today might succeed tomorrow. Maybe if models get good enough, you can just give vague answers and get the results you want, and this lecture is pointless.

The Techniques Stay, The Models Change

Model Evolution

Models get better every month, meaning a tool that fails at a task today might succeed tomorrow. Maybe if models get good enough, you can just give vague answers and get the results you want, and this lecture is pointless.

Engineering Principles

While the tools change, the engineering techniques do not:

1. Writing clear specifications.
2. Providing the right context.
3. Auditing the output.

This lecture covers these skills.

Vibe Coding

What It Looks Like

You give the AI a vague goal ("Make a login page") and run the output.

Vibe Coding

What It Looks Like

You give the AI a vague goal ("Make a login page") and run the output.

What Goes Wrong

- Output changes each time you run the same prompt
- Logic is hard to follow and harder to debug
- You don't fully understand the code you're shipping

English as Code

Prompt as Source

Think of your prompt as source code and the LLM as the compiler.
Ambiguous source means the compiler picks an interpretation, a real compiler would reject the code, but an LLM will make a silent guess (as they are probabilistic models).

English as Code

Prompt as Source

Think of your prompt as source code and the LLM as the compiler.
Ambiguous source means the compiler picks an interpretation, a real compiler would reject the code, but an LLM will make a silent guess (as they are probabilistic models).

Discussion

What does a real compiler do when it hits ambiguous syntax?

English as Code

Prompt as Source

Think of your prompt as source code and the LLM as the compiler.
Ambiguous source means the compiler picks an interpretation, a real compiler would reject the code, but an LLM will make a silent guess (as they are probabilistic models).

Discussion

What does a real compiler do when it hits ambiguous syntax?

Answer

It either rejects the code or picks one parse and runs with it, but LLMs always pick and never say so.

A Better Approach

Specify → Plan → Audit → Implement

Step	What You Do	Output
Specify	Write what the system should do.	`SPEC.md`
Plan	Ask the AI for a step-by-step implementation plan	Plan doc
Audit	Read the plan and flag logical problems	Annotated plan
Implement	AI writes code against the approved plan	Code

A Better Approach

Specify → Plan → Audit → Implement

Step	What You Do	Output
Specify	Write what the system should do.	`SPEC.md`
Plan	Ask the AI for a step-by-step implementation plan	Plan doc
Audit	Read the plan and flag logical problems	Annotated plan
Implement	AI writes code against the approved plan	Code

Why This Order

A bad architecture decision caught in the plan takes 30 seconds to fix. If you implement and then realize the architecture is bad, you have to rewrite everything.

Atomic Tasking

Vague vs. Specific

Vague Prompt	Specific Task
"Fix the sidebar bug."	"1. Open `Sidebar.tsx`. 2. Make `isOpen` persist in `localStorage`. 3. Add a 200ms CSS transition."
"Add search."	"1. Add a `search` query param to `GET /api/posts`. 2. Filter `title` by substring match. 3. Return `[]` (not 404) on no results."

Atomic Tasking

Vague vs. Specific

Vague Prompt	Specific Task
"Fix the sidebar bug."	"1. Open `Sidebar.tsx`. 2. Make `isOpen` persist in `localStorage`. 3. Add a 200ms CSS transition."
"Add search."	"1. Add a `search` query param to `GET /api/posts`. 2. Filter `title` by substring match. 3. Return `[]` (not 404) on no results."

One Task, One Outcome

If you cannot write a single test that passes or fails based on the task, the task is probably too vague. This isn't a hard and fast rule, for simple things a vague prompt can do but once you do more complicated work, it starts to matter more.

Discussion: Writing a Plan

Building a Grading Tool

You are building a Grade Tracker for UMD students.
What should I tell the LLM if I want to build this?

Discussion: Writing a Plan

Building a Grading Tool

You are building a Grade Tracker for UMD students.
What should I tell the LLM if I want to build this?

Data Shape

```
[{ "course": "CMSC132", "credits": 4, "grade": "A" }, ...]
```

Discussion: Writing a Plan

Building a Grading Tool

You are building a Grade Tracker for UMD students.
What should I tell the LLM if I want to build this?

Data Shape

```
[{ "course": "CMSC132", "credits": 4, "grade": "A" }, ...]
```

Answer

1. Map each object: multiply `credits` by its grade-point weight (A=4.0, B=3.0, ...).
2. Sum all products → total quality points. Sum all `credits` → total credits.
3. Divide total quality points by total credits. Round to 2 decimal places.

The Context Window

What It Is

The context window is the LLM's working memory meaning that it holds everything the model can see when it generates a response.

The Context Window

What It Is

The context window is the LLM's working memory meaning that it holds everything the model can see when it generates a response.

Context Rot

When you paste in too much irrelevant code, the model loses track of the actual task and starts making mistakes on things it got right earlier.

The Context Window

What It Is

The context window is the LLM's working memory meaning that it holds everything the model can see when it generates a response.

Context Rot

When you paste in too much irrelevant code, the model loses track of the actual task and starts making mistakes on things it got right earlier.

Fresh Context

When a session gets long and messy, starting a new conversation is often faster than continuing a degraded one.

Token Management

The Budget

Every model has a hard token limit meaning that every token you use costs latency and money. Being very careful about what you load has real cost implications.
As of recent, model providers like OpenAI, Anthropic, etc. have heavily subsidized model inference. However, as the pressure for these companies to actually become profitable increases, their prices will continue to rise. This means that minimizing the tokens you use is in your best interest (usually).

Token Management

The Budget

Every model has a hard token limit meaning that every token you use costs latency and money. Being very careful about what you load has real cost implications.
As of recent, model providers like OpenAI, Anthropic, etc. have heavily subsidized model inference. However, as the pressure for these companies to actually become profitable increases, their prices will continue to rise. This means that minimizing the tokens you use is in your best interest (usually).

The Rule

Load the minimum context that lets the model answer correctly. When it gets something wrong, the missing piece is usually one function or one config value, not the whole codebase.

Explicit vs. Automatic Context

Two Modes

Mode	How It Works
Automatic	Tool indexes your codebase and retrieves what it thinks is relevant
Explicit	You point the tool at a specific file, function, or doc

Explicit vs. Automatic Context

Two Modes

Mode	How It Works
Automatic	Tool indexes your codebase and retrieves what it thinks is relevant
Explicit	You point the tool at a specific file, function, or doc

When the AI Fails

Before blaming the model, check what it could actually see. The root cause is usually missing context, not a bad model.

How AI Actually Reads Your Code

Not Magic, Just CLI Tools

Modern AI CLI agents (like Claude Code or Gemini CLI) don't have magical contextual awareness. They literally run `ls` to see files, `grep` to find keywords, and `cat` to read them. We actually learned about these same tools in class!

How AI Actually Reads Your Code

Not Magic, Just CLI Tools

Modern AI CLI agents (like Claude Code or Gemini CLI) don't have magical contextual awareness. They literally run `ls` to see files, `grep` to find keywords, and `cat` to read them. We actually learned about these same tools in class!

Clean Code Matters More Now

If your code is poorly organized or your variable names are terrible (e.g., `data1`, `temp_val`), `grep` fails. If `grep` fails, the AI fails. Writing clean, modular code has always been important but it also helps out AI (and humans for that matter) if you needed any extra motivation.

MCP: Model Context Protocol

What It Is

MCP is a standard protocol that lets AI agents connect to external tools and data sources.

MCP: Model Context Protocol

What It Is

MCP is a standard protocol that lets AI agents connect to external tools and data sources.

Why It Matters

Without MCP, your AI only knows what you paste into the prompt. With MCP, your AI can run SQL queries against your database, read Jira tickets, and check Slack messages directly.

MCP In Practice

Examples

MCP Server	What It Lets the AI Do
Postgres	Read your schema, run queries
GitHub	Read issues, open PRs
Linear	Update tickets as work completes
Slack	Post to a channel on deploy

MCP In Practice

Examples

MCP Server	What It Lets the AI Do
Postgres	Read your schema, run queries
GitHub	Read issues, open PRs
Linear	Update tickets as work completes
Slack	Post to a channel on deploy

The Result

Instead of copy-pasting database output into the chat, the AI queries the database directly and sees live data.

Your New Job

From Author to Reviewer

Prompting AI for code shifts your role from author to reviewer. The AI opens the pull request, and you decide whether it merges.

Your New Job

From Author to Reviewer

Prompting AI for code shifts your role from author to reviewer. The AI opens the pull request, and you decide whether it merges.

Accountability

You own everything you ship, regardless of who wrote it. "The AI generated it" is not a defense when a bug reaches production.

AI-Generated Code and Security

The Security Gap

AI-coauthored pull requests contain 2.74x more security vulnerabilities than human-only PRs, with XSS, logic flaws, and missing input validation being the most common. [1]

[1] CodeRabbit, State of AI vs Human Code Generation Report, December 2025

AI-Generated Code and Security

The Security Gap

AI-coauthored pull requests contain 2.74x more security vulnerabilities than human-only PRs, with XSS, logic flaws, and missing input validation being the most common. [1]

[1] CodeRabbit, State of AI vs Human Code Generation Report, December 2025

Read More, Write Less

Generating code is the easy part but reading and auditing the output requires more effort.

Code Auditing

Three Practices

Practice	What You Do
Plan Audit	Read the full plan before the AI writes any code
Red-Teaming	Ask the AI to critique its own output: "List 3 potential race conditions in this c
TDD-First	Write tests before implementation so the AI codes against your spec

Debugging with AI

The "Wrong" Way

Pasting a 50-line stack trace and saying "Fix this."

Debugging with AI

The "Wrong" Way

Pasting a 50-line stack trace and saying "Fix this."

The "Right" Way

Provide the AI with:

1. The error trace
2. The specific environment (e.g., "Node 18, React Native 0.72")
3. The expected behavior

Debugging with AI

The "Wrong" Way

Pasting a 50-line stack trace and saying "Fix this."

The "Right" Way

Provide the AI with:

1. The error trace
2. The specific environment (e.g., "Node 18, React Native 0.72")
3. The expected behavior

Why This Matters

AI is eager to please. If you just say "fix this," it might delete the feature entirely to make the error go away, rather than fixing the underlying logic.

Trust But Verify

What "Done" Means

AI Says	What You Do
"I've updated the database schema."	Check that migration files were generated.
"The bug is fixed."	Run `npm test` and confirm the failing case now passes.
"This code is secure."	Ask: "How would an attacker bypass this `if` statement?

Trust But Verify

What "Done" Means

AI Says	What You Do
"I've updated the database schema."	Check that migration files were generated.
"The bug is fixed."	Run `npm test` and confirm the failing case now passes.
"This code is secure."	Ask: "How would an attacker bypass this `if` statement?

The Rule

Treat every AI claim as something to verify, not something to assume.

What We Covered

Four Areas

Area	Idea
Productivity	Net time gain, not line count; experience changes the calculus
Specification	Write explicit, verifiable tasks before prompting, and commit first
Context	Code must be clean for AI to search it; load the minimum context
Code Review	Treat every AI PR as something to review, not assume

What We Covered

Four Areas

Area	Idea
Productivity	Net time gain, not line count; experience changes the calculus
Specification	Write explicit, verifiable tasks before prompting, and commit first
Context	Code must be clean for AI to search it; load the minimum context
Code Review	Treat every AI PR as something to review, not assume

Tools

Cursor, Windsurf, Claude Code, and MCP servers are all interfaces for these ideas. Once you understand the underlying primitives, the tools become much easier to use well.

This Is Our Last Class!

Thank you all for showing up to and participating in class. I've had a lot of fun teaching this class and I hope you have learned something that is actually useful in your day to day life or careers. Some parting asks, and unsolicited advice:

1. Connect with me on LinkedIn! I'd like to see what you all get up to in the future: www.linkedin.com/in/mohammad-durrani/
2. If you ever need any help with internships / jobs / interviews, I am always happy to help and to be a resource.
3. Note: Using your judgement and building taste in software engineering. Be opinionated!
4. Note: Keeping your brain sharp.
5. Spread the knowledge of the tools that you use to your coworkers, classmates, friends etc.
6. Fill out the course evaluation! I want to continue teaching in the future and would love to know what I can be better at. You will also get some extra credit!

Best of luck with everything that you do in the future!