From Myth to Legend

How Generative AI can Supercharge Productivity to Create 10x Developers

Gautam Korlam
Principal Engineer

Serdar Badem
Lead Product Manager
Myth of the 10X developer
100% Knowledgeable

100% Productive

100% Perfectionist
Modern software complexity

Thousands of services and hundreds of thousands of RPC calls
How Standards Proliferate:
(See: A/C chargers, character encodings, instant messaging, etc.)

Situation: There are 14 competing standards.

14?! Ridiculous! We need to develop one universal standard that covers everyone’s use cases. Yeah!

Soon:

Situation: There are 15 competing standards.
Enter Generative AI

Analysts’ opinion of where the GenAI hype cycle is

Where we are today

Innovation trigger | Peak of inflated expectations | Trough of disillusionment | Slope of enlightenment | Plateau of productivity

GenAI Hype Cycle
Expectations vs (alternate) reality

function scheduleProcessor() { 

Hello, this is a ride share app that allows users to create an account to book a ride. The ride will be stored in a database and can be viewed by others. The user can also view other rides and request to join them. The user can also view their own rides and delete them if they wish. The user can also view their own profile and edit it if they wish. The user can also view other users profiles and view their rides. The user can also view their own rides and delete them if they wish. The user can also view their own profile and edit it if they wish. The user can also view other users profiles and view their rides. The user can also view their own rides and delete them if they wish. The user can also view other users profiles and view their rides. The user can also view their own rides and delete them if they wish. The user can also view other users profiles and view their rides. The user can also view their own rides and delete them if they wish. The user can also view other users profiles and view their rides. The user can also view their own rides and delete them if they wish. The user can also view other users profiles and view their rides. The user can also view their own rides and delete them if they wish. The user can also view other users profiles and view their rides. The user can also view their own rides and delete them if they wish. The user can also view other users profiles and view their rides. The user can also view their own rides and delete them if they wish. The user can also view other users profiles and view their rides. The user can also view their own rides and delete them if they wish. The user can also view other users profiles and view their rides. The user can also view their own rides and delete them if they wish. The user can also view other users profiles and view their rides. The user can also view their own rides and delete them if they wish. The user can also view other users profiles and view their rides. The user can also view their own rides and delete them if they wish.
Gen AI opportunities in SDLC
AI opportunities in the developer workflow

**Design**
- Documentation
- Design Review
- UX Design
- Knowledge Base, Support
- Tailored to your enterprise

**Develop**
- Code Generation
- Code Quality
- Automated Fixes

**Release**
- End to End Testing

**Manage**
- Incident Management
Automatic documentation freshness

Code Commits
- Generate meaningful docs from commit descriptions

Chat messages
- Capture tribal knowledge from conversations in slack

Update Docs
- Search for relevant docs and update with distilled information

Search for relevant docs and update with distilled information
Automatic review of design requirements

- Capture comprehensive requirements first time
- Review design documents, provide automatic suggestions for improvement
- Suggest changes to design based on existing user comments

PRD

PM, Product Ops, Sec, Privacy, Legal

ERD

Developer, Eng team, Eng Sec, Privacy

Developer, Eng team
Faster UX design iteration

```javascript
import React from 'react';
import { Button } from 'baseui/button';
import { Input } from 'baseui/input';

const FruitButtons = () => {
  return (
    <Button>Apple</Button>
    <Button>Orange</Button>
    <Button>Grape</Button>
    <Input placeholder="Type here..." />
  );
};

export default FruitButtons;
```
Our experience using code assistants

Internal survey results

- Effective for simple boilerplate code and tasks
  Copilot is really useful for writing 1-2 lines of straightforward / repetitive code

- Code suggestions frequently need further editing
  It frequently has small misses. I find myself usually double-checking Copilot's work for longer than it would have taken me to just write that work myself.

29% Acceptance rate
1.4 LOC per accepted code
75% Feel more productive
6.0 NPS score for Copilot
Code assistants are useful for simpler tasks
Gen AI beyond code completion
Uber code assistant

IDE context

Code suggestions

Fine tuned LLM

Based on open-source Codellama

Uber code base

IDE

Linux

JavaScript

Go

Java

Python

Apple
Automatically improve code quality

- Suggest fixes in the IDE & local builds
  - Code corrections while writing it
    - Auto-fix suggestions within the IDE

- Fix new errors in PRs
  - Write initial code
  - Review & approve changes
  - Suggest an automatic fix

- Continuously fix tech debt in the existing code base
  - Generate Pull requests with fixes
  - Review & approve changes
Impact of errors in code

Incident in 2021 due to MisusedWeekYear

“Week year” is intended to be used for week dates, e.g. “2015-W01-1”, but is often mistakenly used for calendar dates, e.g. 2014-12-29

```java
@SuppressWarnings("MisusedWeekYear")
private static final DateTimeFormatter IDEMPOTENCY_DF=
DateTimeFormatter.ofPattern("MM_YYYY");
```
Automatically fix errors in code

Code analysis

Large Language Model

AI Prompt Engineering

Fixed Code

3.5K
Errors auto-fixed / week*

875
Dev hrs saved per week*
Automatically fix CI errors

Pipelines > iOS Build
Test
Build #4 master HEAD
Job ...
Job ...

Develop

Cl build
Local development
Code review
Submit Queue
Continuous Delivery

CI Failure
Rate is HIGH
CI execution Time is HIGH

Build or test failure
Changes requested
Build, test or merge failure
Build failure

Root cause, reproduce and debug

HIGH IMPACT WHEN
Automatically fix code review feedback

- **10K** Diffs / week
- **1hr** Dev time/diff
- **~$1M** Cost per week
- **2K** Auto-fixes / week
- **2K** Dev hrs saved per week
- **$10M** Savings per year

Reviewer comment

```java
byte[] serialize = thriftJsonSerializer.serialize(mapFeature);
String mapFeatureAsString = objectMapper.writeValueAsString(serialize);
LOGGER.info("MapFeature being written as '{}', mapFeatureAsString);
```

Maybe turn these into debug logs... can pollute heavily

AI suggested fix

```java
@893 ObjectMapper objectMapper = new ObjectMapper();
   - LOGGER.info("Pushing auto-applied events to heatmap for {}",
     getHceRequestContext().getUsId());
   - LOGGER.debug("Pushing auto-applied events to heatmap for {}",
     getHceRequestContext().getUsId());
```

Categories of review feedback fixed

- Rename, annotate or change type (11%)
- Nitpicks (3%)
- Other actionable feedback (6%)
  - Logging
  - Exception handling
  - Null checks
  - Collection & streams
  - Date & time APIs
  - Missing documentation
Automatically fix issues in the IDE

Implicit capture-by-reference of free variables in goroutines

```go
for _, item := range items {
    go func() {
        // item
    }()
}
```

Concurrent accesses to Go's built-in, thread-unsafe maps

```go
resultMap := map[UID]string {}
for _, uuid := range UUIDs {
    go func(key) {
        if ...
            resultMap[key] = Process(key)
    } (uuid)
}
```

Mixing message-passing with shared-memory

```go
func (f *Future) Start() {
    go func() {
        f.result = f.fPtr()
        f.ch <- 1
    }()
}

func (f *Future) Wait(ctx Context) error {
    select {
    case <- f.ch:
        return nil
    case <- ctx.Done():
        f.result = TimeOutErr
    }
    return TimeOutErr
}
```

Concurrent accesses to Go's built-in, thread-unsafe maps cause frequent data races

View Problem (F8)  Accept Fix

✨ AI suggested fix for data races
Automatic app testing

AI generated mobile test flow

Eliminates costly test maintenance

Higher quality, less effort
Efficient incident management

**On-Call**
- Pager-Duty
- Rotation, alerts, runbooks
- Improve signal to noise ratio

**Incident response**
- Identify incident and respond
- Mitigate incident
- Communicate incident

**Postmortem**
- Identify root cause
- Write post-mortem
- Peer review
- Select incidents for review
- Conduct incident review
- Plan and prioritize follow-up work

**Incident review**
- Find frequent fault areas
- Create follow up actions

- Al generated Handoff notes
- Anomaly detection Incident leveling
- Enrich with similar incidents
- Catch me up with status
- Find the root cause
- Identify related events
- Suggest mitigation steps
- Generate post mortem
- Assess the post mortem
Automate developer support

Slack bot to answer questions

**gkorlam** 1 minute ago
!obs how do I setup a new alert?
1 reply

**BMO** 1 minute ago
Thinking... Genie 🤖 is contemplating your 👩‍💻 query...

**gkorlam** 1 minute ago
Iobs how do I setup a new alert?

**BMO** 1 minute ago
Genie is still in beta. Please read [http://t.uber.com/genie-lm](http://t.uber.com/genie-lm) for more details.

Setting up a new alert in the H2 system can be easily done within the uMonitor interface. Here are the steps:

1. Go to [uMonitor](https://umonitor.uberinternal.com/).
2. Create a new alert group if one doesn’t already exist for your service. To do this:
   1. Click **[New Alert Group]**
   2. Enter the alert group name.
   3. Enter a team or individual email address in the description.

Show more
The Future
How to get started?
Considerations

- **Which model?**
  - Out-of-box LLMs via APIs
  - Tune LLMs via APIs
  - Vendor hosted (OpenAI, Google Vertex, Antrophic)
  - VPC (Generic knowledge Company Limited)
  - On-prem (Company High context)

- **Where to host**
  - Vendor hosted
  - VPC
  - On-prem

- **Which LLM**
  - OpenAI
  - Google Vertex
  - Antrophic
  - Generic knowledge Company Limited
  - Company High context

- **Knowledge context?**
  - Critical
  - High
  - Medium
  - Low

- **Time to market**
  - Critical
  - High
  - Medium
  - Low

- **Cost**
  - $$ $$ $$ $$

- **Latency**
  - Sensitive
  - High
  - Medium
  - Low
  - Offline
Evaluate potential risks

**RISKS**

<table>
<thead>
<tr>
<th>Data sent to LLMs</th>
</tr>
</thead>
<tbody>
<tr>
<td>● <strong>Data privacy</strong> - Sharing private data with a 3rd party</td>
</tr>
<tr>
<td>● <strong>IP</strong> - Risk of exposing own IP into public domain</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Data suggestions from LLM</th>
</tr>
</thead>
<tbody>
<tr>
<td>● <strong>IP</strong> - Risk of getting back other’s copyrighted content</td>
</tr>
<tr>
<td>● <strong>Quality</strong> - Accuracy of the outputs</td>
</tr>
<tr>
<td>● <strong>Explainability</strong> - Not being able to trace back the output to its source</td>
</tr>
</tbody>
</table>
Get your team excited

Internal Hackathon

- 700+ Participants
- 100 Projects demoed
- 6 Sites around the globe
- 3 Broad categories

- Product experience
- Developer productivity
- Business operations
Takeaways

- Modern software development is complex
- Gen AI promises to tame this complexity
- Lots of opportunities in realizing these promises
- The entire Developer workflow needs to be rethought with AI
- Let’s get started!
Thank you!

Questions?
Backup
Developer workflow: AI opportunities

**Design**
- Code & test generation
- Automated fixes
- Code review

**Develop**
- Requirements
  - Write requirements
  - Review design docs
  - Risk assessment
- UX design
  - UI/UX design

**Release**
- Testing
- Automated E2E & mobile testing
  - Auto generate mobile, localization tests

**Manage**
- AI Ops
- Incident mitigation:
  - Troubleshoot analytics events
  - On-call assistant

**Knowledge Base (Messaging, documentation, collaboration tools)**
- Documentation: Auto generate & update documentation
- On-call experience: Chatbot for internal knowledge base
Design ERD automatic review

Prompt Engineering
Few-shot
Chain of thoughts
Self debugging
Tree of thoughts

Reviewed design document

User review / acceptance

Generate Code/Test Fix

VectorDB

Design Documentation

Feedback & comments
Automatically fixing static analysis errors

Incident in 2021 due to MisusedWeekYear

```java
@SuppressWarnings("MisusedWeekYear")
private static final DateTimeFormatter IDEMPOTENCY_DF=
DateTimeFormatter.ofPattern("MM_YYYY");
```

“Week year” is intended to be used for week dates, e.g. “2015-W01-1”, but is often mistakenly used for calendar dates, e.g. 2014-12-29

- **7K** Java errors / week
- **15 min** Dev time / Error
- **3.5K** Errors auto-fixed / week
- **875** Dev hrs saved per week

10K @SuppressWarnings in Java codebase

Other analyses in progress
- Go data races

* Google study

---

Sorry, that extra $35 you got in your Uber account was a mistake.
AI-driven automated Mobile End-2-End Testing

E2E automation challenges

- Flakiness
- Maintenance
- Different languages
- Impossible to shift left

AI-based E2E testing
Uber Tech-wide HackDayz

713 Participants
98 Projects demoed
21 Finalists selected

6 Sites around the globe
3 Broad categories
6 Winners

Developer productivity
Code generation, Code reviews, Automated fixes, Testing & debugging, Knowledge base

Business operations
Operations improvements, Data & query optimization, Security, Privacy & safety risk, Employee experience

Product experience
Delivery experience, Rider experience, Earner experience
How to create internal LLM
What we have been able to get it to do

Design
- PRD
- ERD

Develop
- Local development
- CI build
- Code review
- Submit Queue
- Continuous Delivery

Release
- Rollout

Manage
- Production management

Coding assistant
- Copilot Live
- Copilot X Upcoming

Automated fixes
- Static / dynamic analysis
- Review comments
- Code migrations

Automated testing
- Automated mobile E2E testing

Test failures
- Bugs
- Service rollbacks
- Monitoring
- Crashes

Incidents
- Monitoring
- Bugs
- Crashes
Gen AI or Not?

👍 Great if your task is ...

- Related to natural language processing or image/video/audio creation
- Translation, summarization, sentiment analysis, extraction, search, similarity, retrieval
- Requires general knowledge and/or reasoning
- You have very little training data
  - i.e., zero / few shot learning

👎 Not so great if ...

- you already have an effective ML solution that’s performing well
- (e.g., NLP classification)
- The cost of being inaccurate IS very high
- Explainability of a prediction IS very important
- Your use case is highly sensitive to latency
Our color palette embraces a full spectrum of color and has been brightened for its legibility and accessibility. All colors are ADA compliant when used as a background paired with black text or when used as colored typography paired with a black background.

For **Mobility**, be sure to use white or color headline text on a black background. For **Delivery**, be sure to use black text over a color background.
Let’s get started!