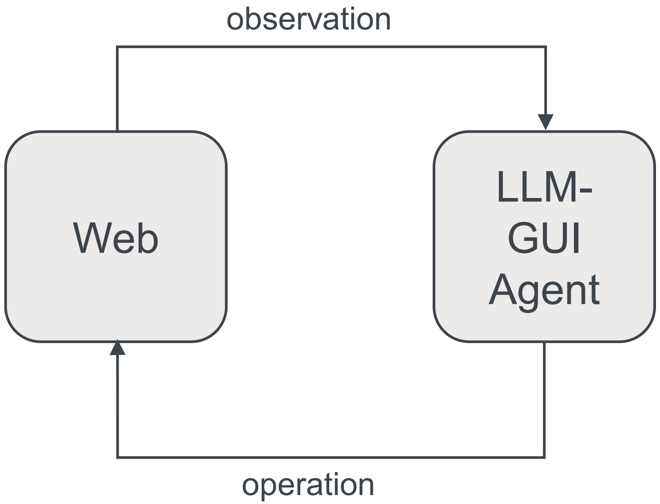
University of Stuttgart
Institute of Industrial Automation
and Software Engineering



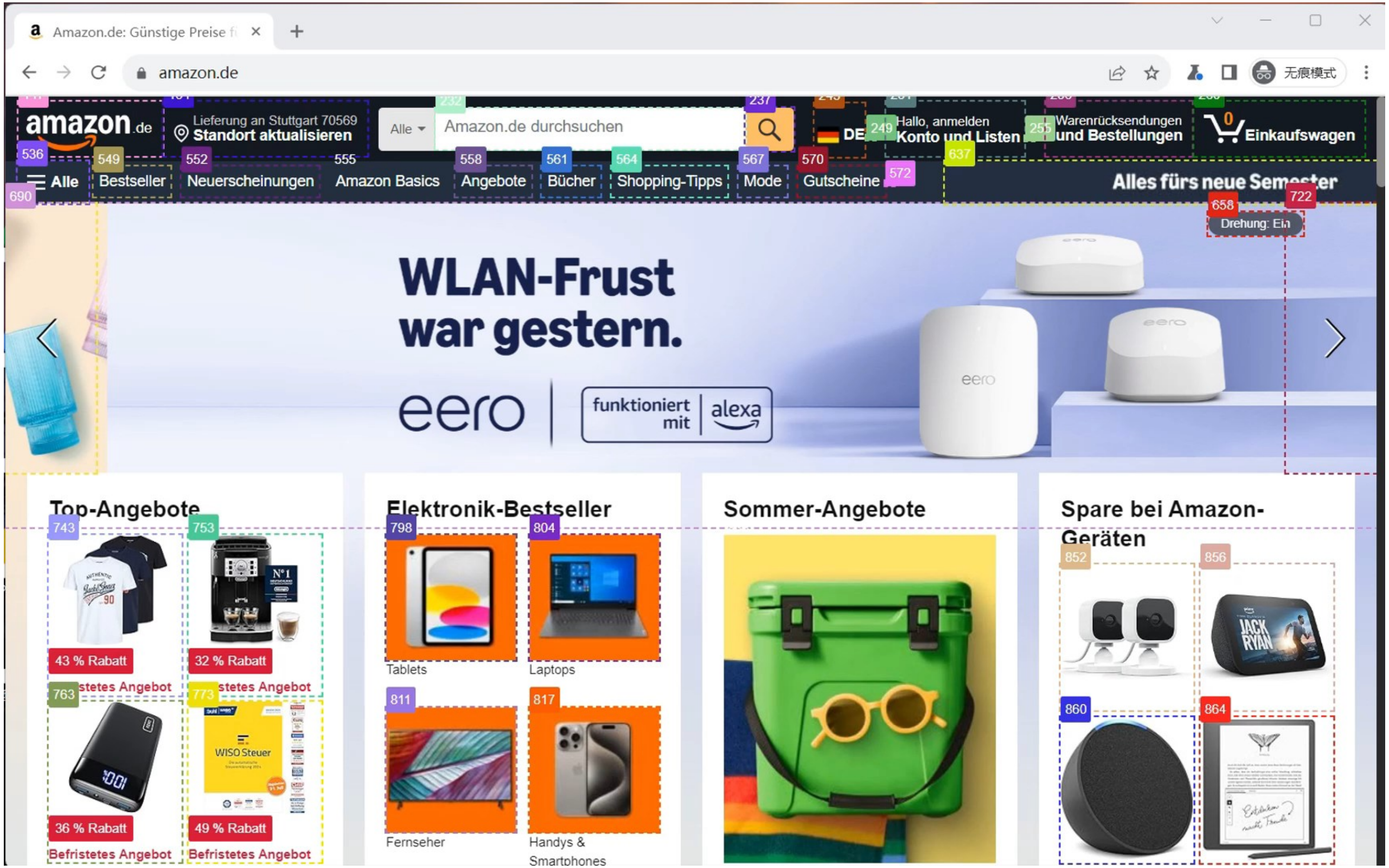
Design and Implementation of an LLM-Powered GUI Agent System for Automated Interface Interaction

Zhe CAO
Elektromobilität
Supervisor: Yuchen Xia M.Sc.

Quick Demo

LLM Agent: Shopping for the user

- User input: Go to Amazon, find a book called ‘Build a Large Language Model from Scratch’ and add it to the cart.



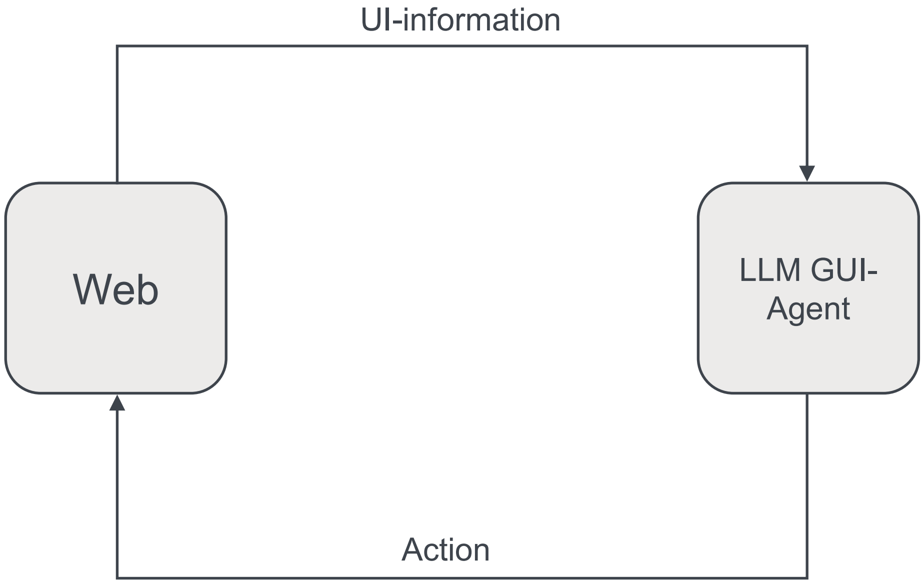
Agenda

1. Problem statement and use case
2. Background
3. Basics
4. System design
5. Test and evaluation
6. Conclusion and outlook

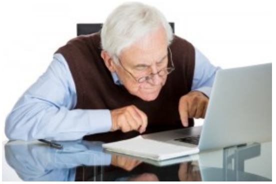
Problem statement

Problem statement & Use case

Leveraging LLMs and VLMs for Human-Computer Interaction



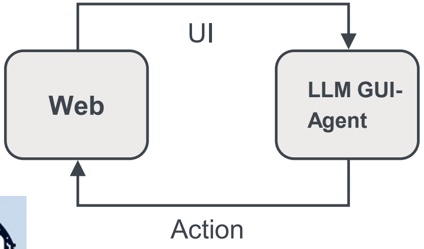
- **Easier human-computer interaction**
  - making it easier for **people with low computer skill**
  - and **disabled people** to use computers to do what they want through the browser.



Background

Background

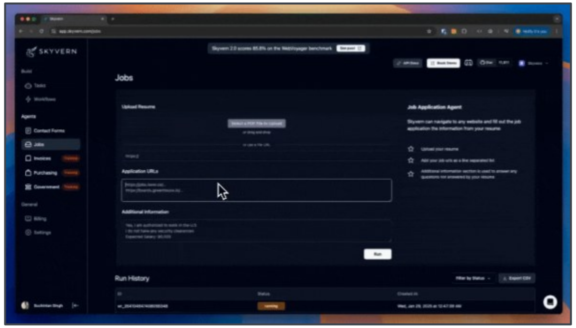
Related work(until 03.2025)



OpenAI Operator

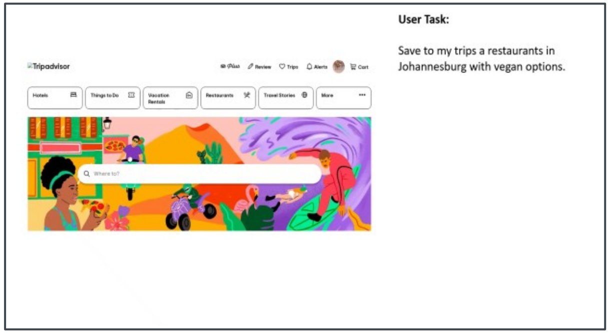
launched in 02.2025

SKYVERN



launched in 03.2025

OmniParser



User Task:

Save to my trips a restaurants in Johannesburg with vegan options.

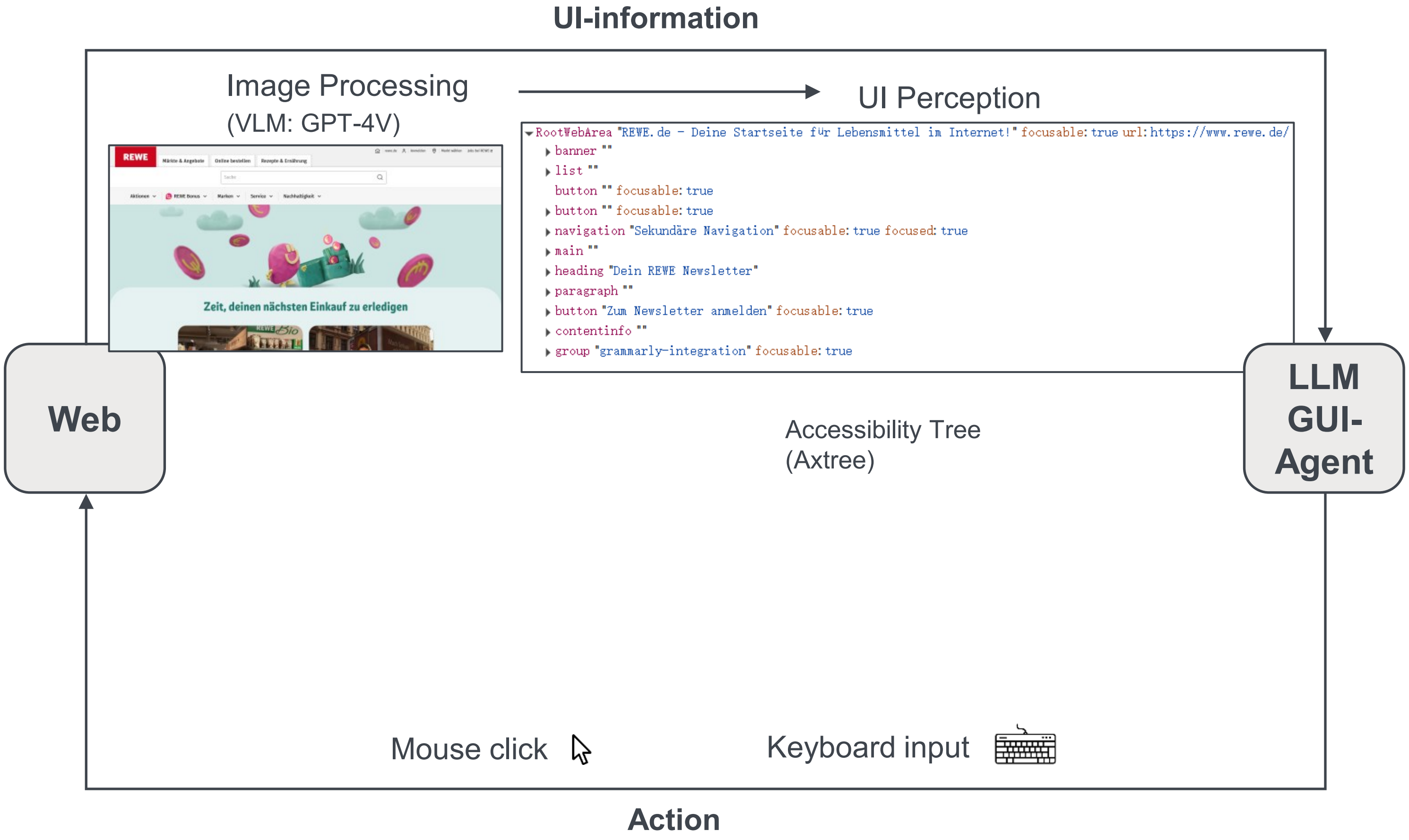
launched in 03.2025

- Master thesis contributions(begin from 02.2025)
  - Building a system with open-source tools to reverse-engineer a similar result without knowing the exact implementation details.
  - Evaluation & examinate the limitation.

Basics

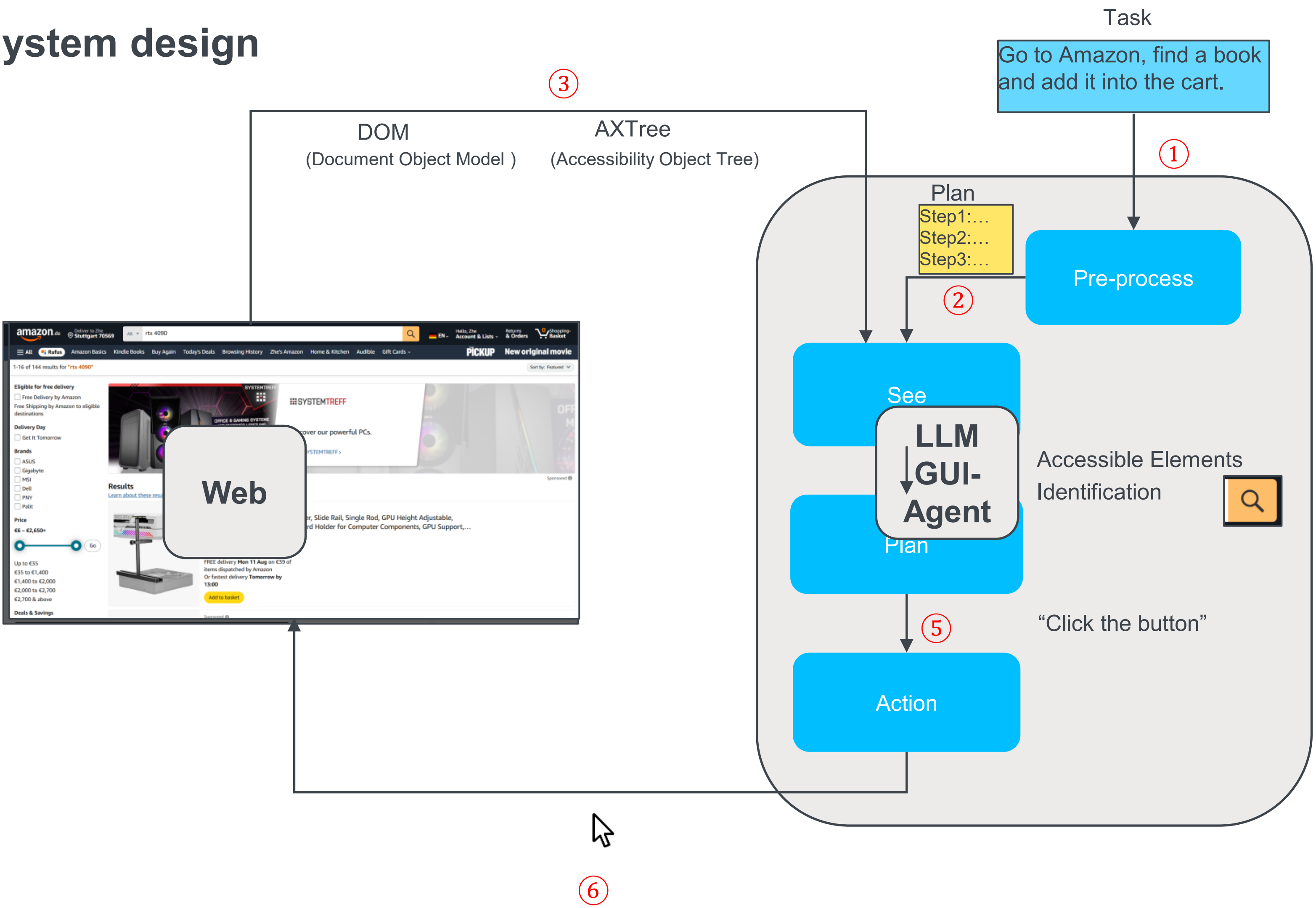
Basics

Framework



```
▾RootWebArea "REWE.de – Deine Startseite für Lebensmittel im Internet!" focusable: true url: https://www.rewe.de/
  ▸banner ""
  ▸list ""
   button "" focusable: true
  ▸button "" focusable: true
  ▸navigation "Sekundäre Navigation" focusable: true focused: true
  ▸main ""
  ▸heading "Dein REWE Newsletter"
  ▸paragraph ""
  ▸button "Zum Newsletter anmelden" focusable: true
  ▸contentinfo ""
  ▸group "grammarly-integration" focusable: true
```

System design



Task

Go to Amazon, find a book and add it into the cart.

①

Plan

Step1:…
Step2:…
Step3:…

Pre-process

②

③

DOM
(Document Object Model)

AXTree
(Accessibility Object Tree)

See

LLM GUI-Agent

Plan

Accessible Elements Identification

“Click the button”

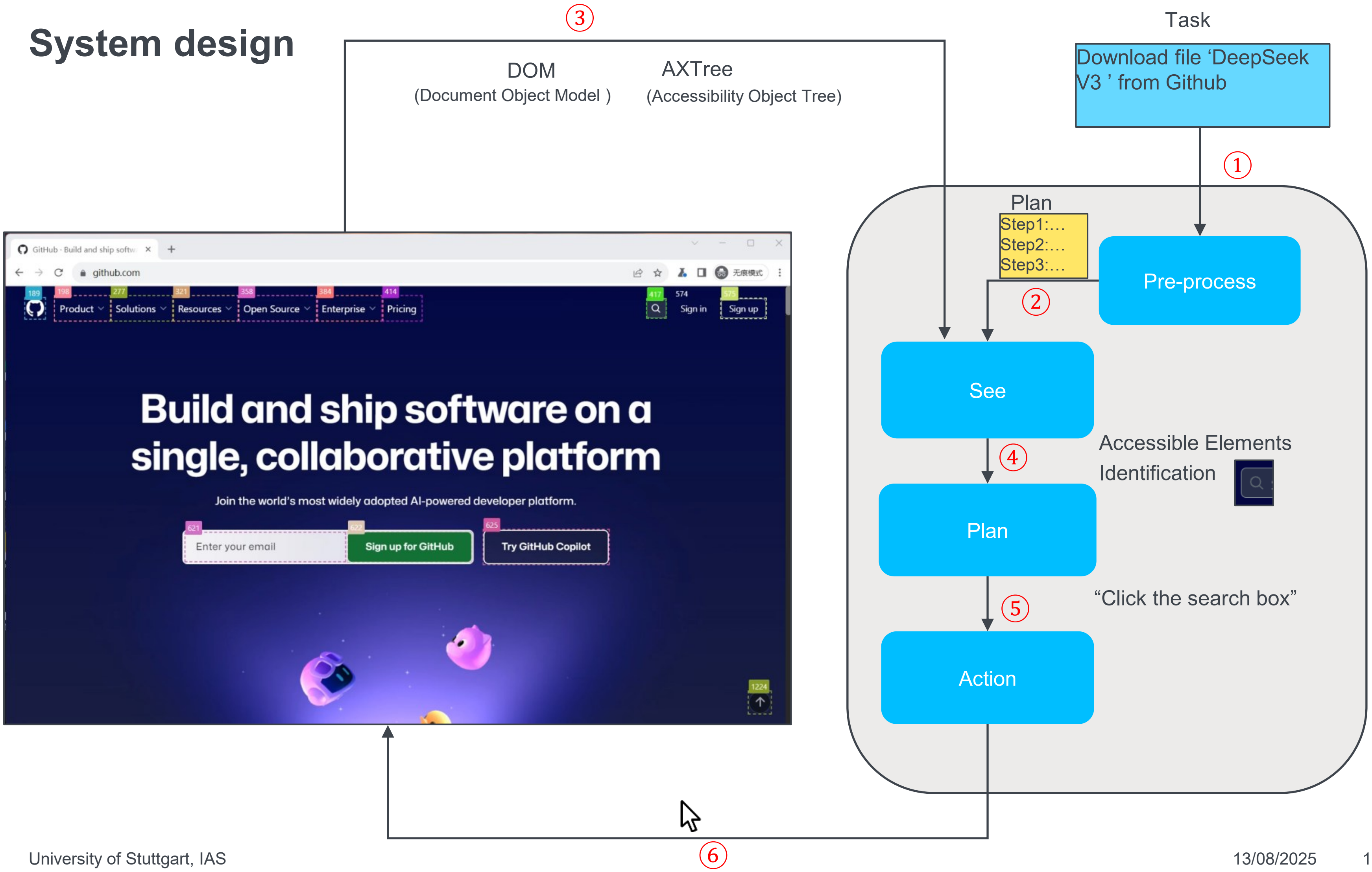
⑤

Action

Web

⑥

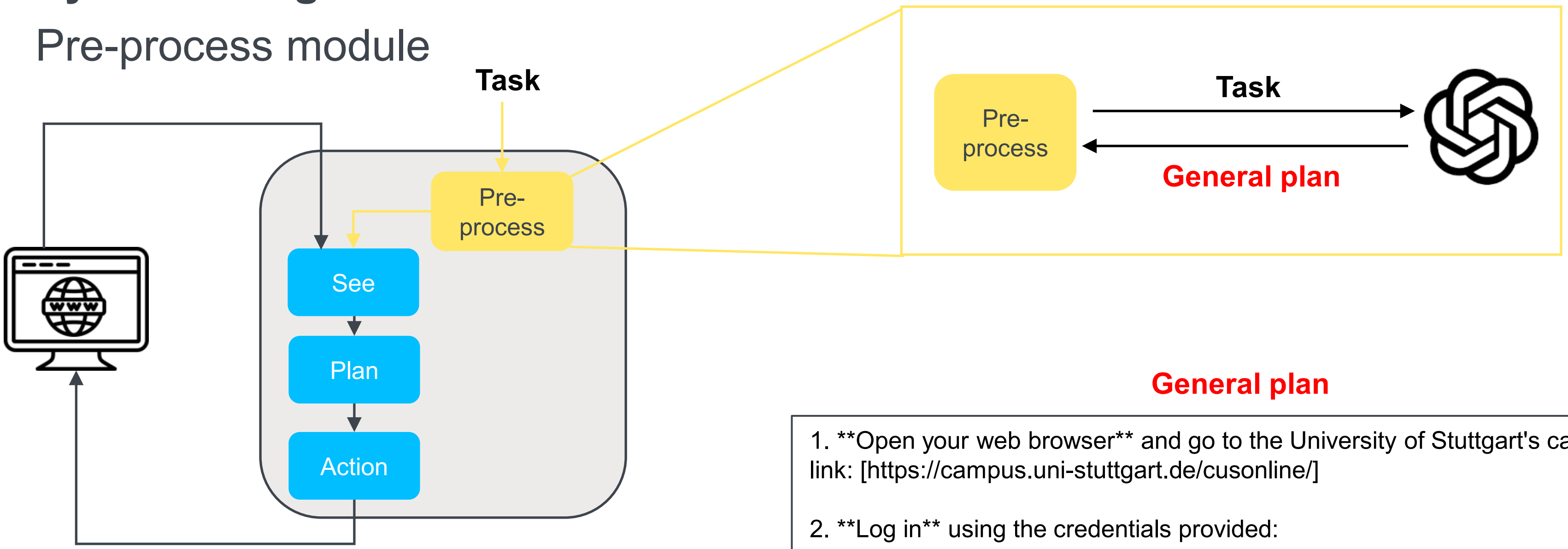
System design


③
DOM
(Document Object Model)
AXTree
(Accessibility Object Tree)
Task
Download file 'DeepSeek V3 ' from Github
①
Plan
Step1:…
Step2:…
Step3:…
Pre-process
②
See
④
Accessible Elements Identification
Plan
"Click the search box"
⑤
Action
⑥
GitHub · Build and ship softw
github.com
Product
Solutions
Resources
Open Source
Enterprise
Pricing
Sign in
Sign up
Build and ship software on a single, collaborative platform
Join the world's most widely adopted AI-powered developer platform.
Enter your email
Sign up for GitHub
Try GitHub Copilot

System design

System design

Pre-process module



Task

Goal: Login in the university Stuttgart campus system and select a course called Automatisierungstechnik I with username:… and password:…

Website: https://www.google.com/

General plan

1. **Open your web browser** and go to the University of Stuttgart's campus link: [https://campus.uni-stuttgart.de/cusonline/]

2. **Log in** using the credentials provided:
   - Username: `…`
   - Password: `…`

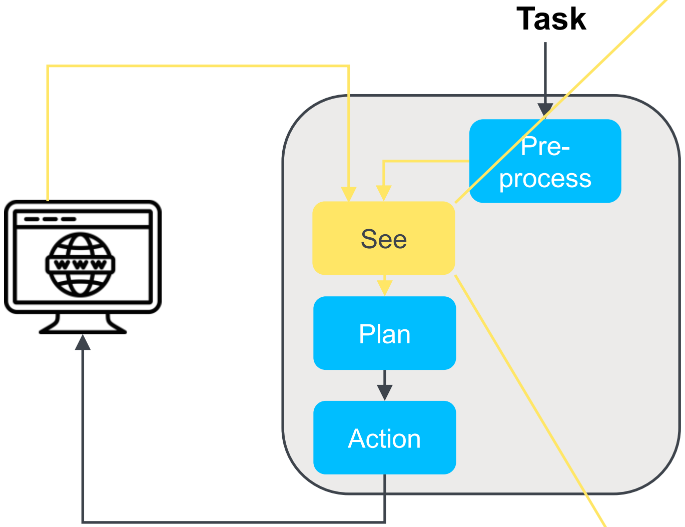
3. Once logged in, **search for the course** titled “Automatisierungstechnik I - Vorlesung” in the course catalog or dashboard.

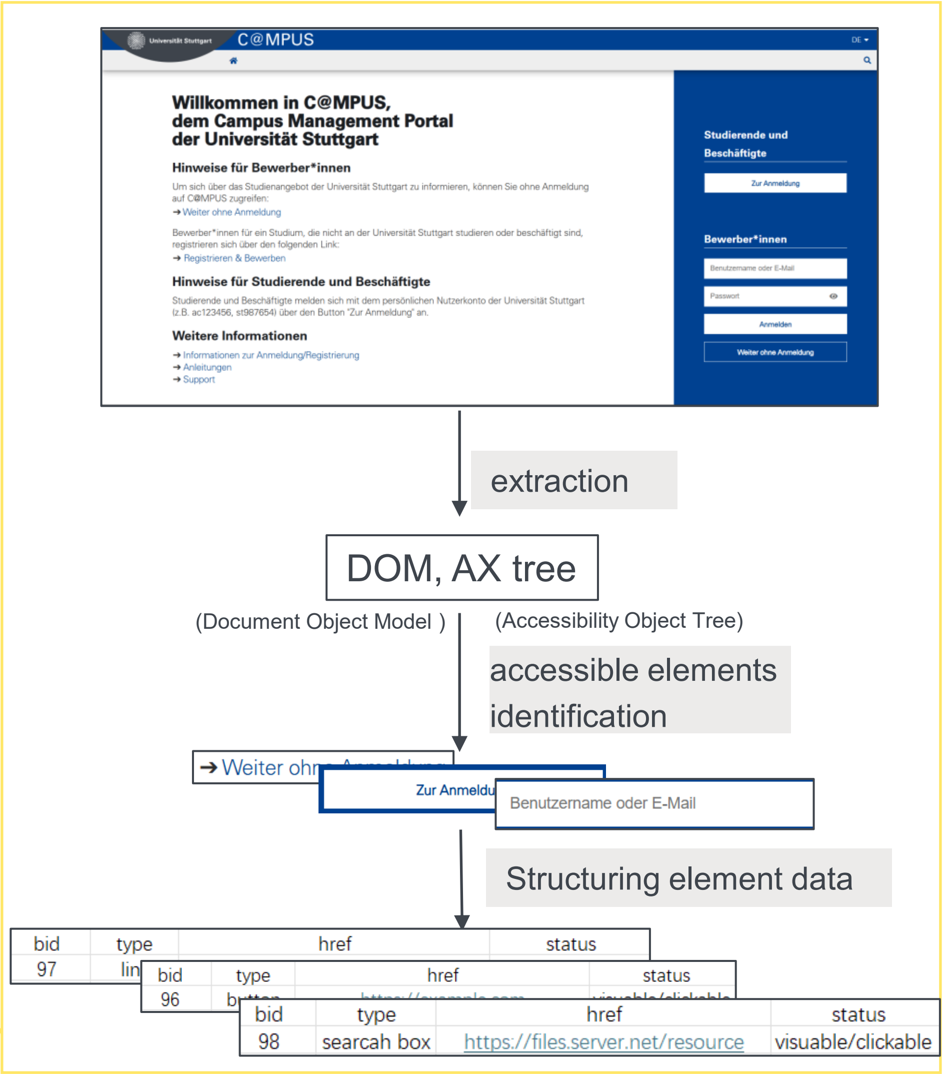
4. Click on the course link to access the course materials, schedule, and any other relevant information.

5. If you encounter any issues logging in or finding the course, ensure your credentials are correct or check for any university announcements regarding system maintenance.

System design

See module





C@MPUS

DE

Willkommen in C@MPUS, dem Campus Management Portal der Universität Stuttgart

Hinweise für Bewerber*innen

Um sich über das Studienangebot der Universität Stuttgart zu informieren, können Sie ohne Anmeldung auf C@MPUS zugreifen:

➔ Weiter ohne Anmeldung

Bewerber*innen für ein Studium, die nicht an der Universität Stuttgart studieren oder beschäftigt sind, registrieren sich über den folgenden Link:

➔ Registrieren & Bewerben

Hinweise für Studierende und Beschäftigte

Studierende und Beschäftigte melden sich mit dem persönlichen Nutzerkonto der Universität Stuttgart (z.B. ac123456, st987654) über den Button "Zur Anmeldung" an.

Weitere Informationen

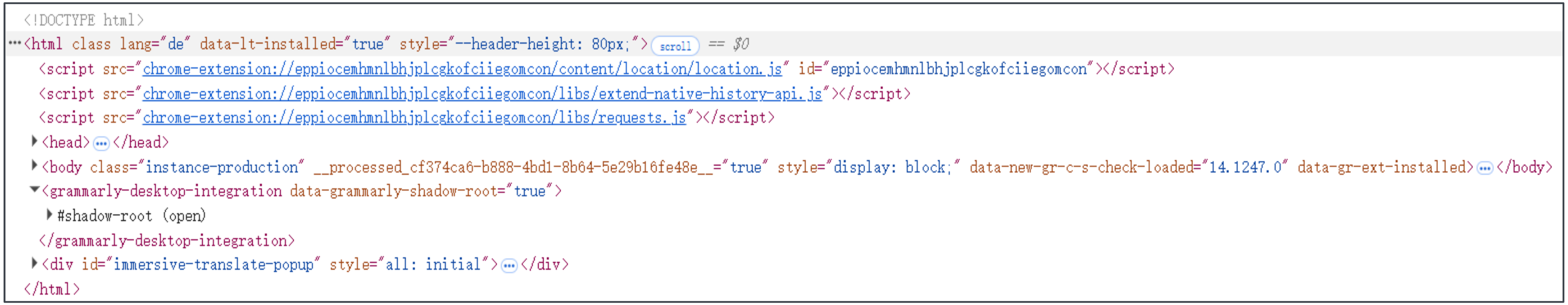
➔ Informationen zur Anmeldung/Registrierung
➔ Anleitungen
➔ Support

bid	type	href	status
97	lin		

bid	type	href	status
96	button	https://example.com	visuable/clickable

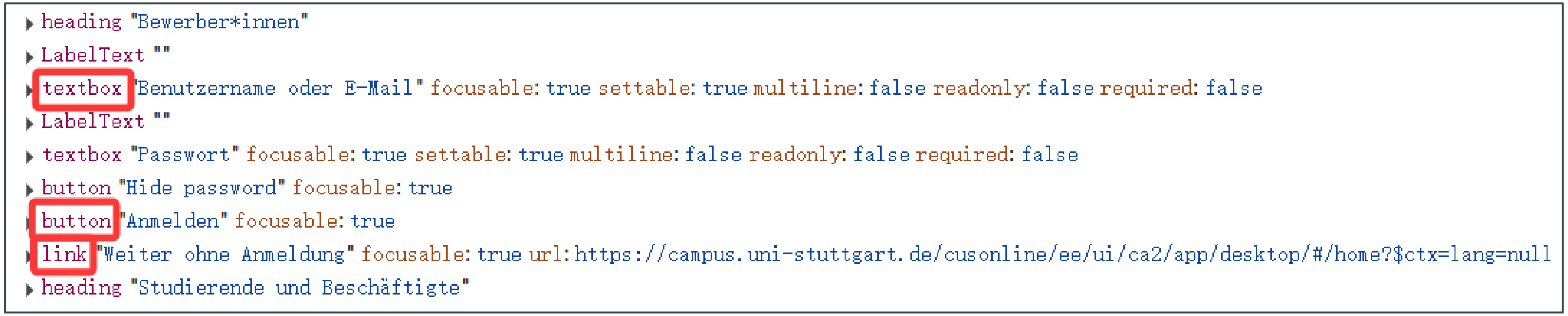
bid	type	href	status
98	searcah box	https://files.server.net/resource	visuable/clickable

DOM (Document Object Model)



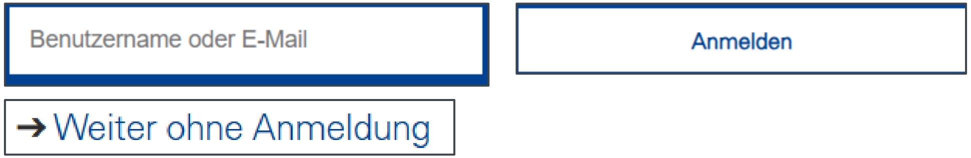
```
<!DOCTYPE html>
···<html class lang="de" data-lt-installed="true" style="--header-height: 80px;"> scroll == $0
   <script src="chrome-extension://eppiocemhmnlbhjplcgkofciiegomcon/content/location/location.js" id="eppiocemhmnlbhjplcgkofciiegomcon"></script>
   <script src="chrome-extension://eppiocemhmnlbhjplcgkofciiegomcon/libs/extend-native-history-api.js"></script>
   <script src="chrome-extension://eppiocemhmnlbhjplcgkofciiegomcon/libs/requests.js"></script>
  ▸<head>···</head>
  ▸<body class="instance-production" __processed_cf374ca6-b888-4bd1-8b64-5e29b16fe48e__="true" style="display: block;" data-new-gr-c-s-check-loaded="14.1247.0" data-gr-ext-installed>···</body>
  ▾<grammarly-desktop-integration data-grammarly-shadow-root="true">
    ▸#shadow-root (open)
   </grammarly-desktop-integration>
  ▸<div id="immersive-translate-popup" style="all: initial">···</div>
 </html>
```

AX tree (Accessibility Object Tree)

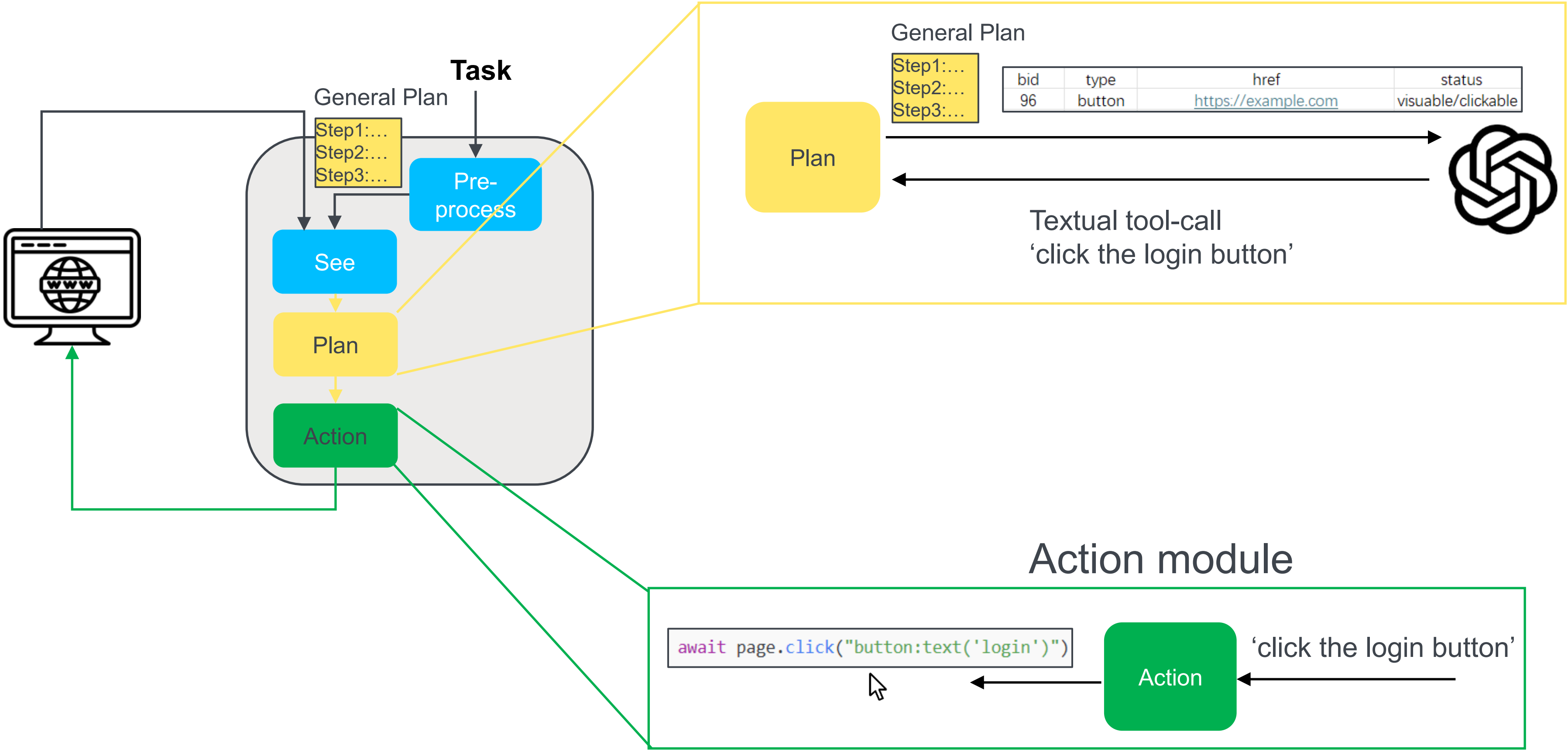


```
▸ heading "Bewerber*innen"
▸ LabelText ""
▸ textbox "Benutzername oder E-Mail" focusable: true settable: true multiline: false readonly: false required: false
▸ LabelText ""
▸ textbox "Passwort" focusable: true settable: true multiline: false readonly: false required: false
▸ button "Hide password" focusable: true
▸ button "Anmelden" focusable: true
▸ link "Weiter ohne Anmeldung" focusable: true url: https://campus.uni-stuttgart.de/cusonline/ee/ui/ca2/app/desktop/#/home?$ctx=lang=null
▸ heading "Studierende und Beschäftigte"
```

interactive elements



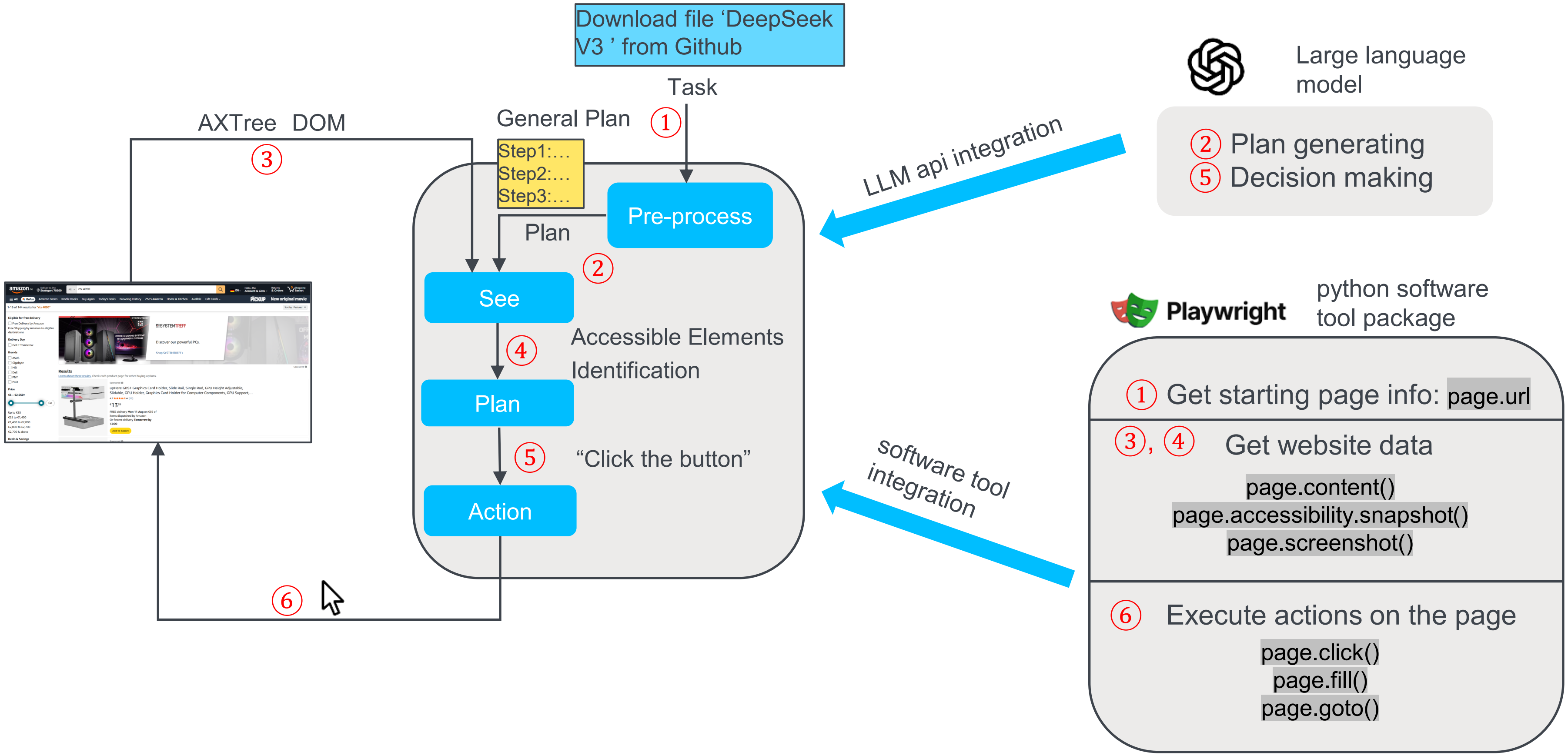
System design



bid	type	href	status
96	button	https://example.com	visuable/clickable

```
await page.click("button:text('login')")
```

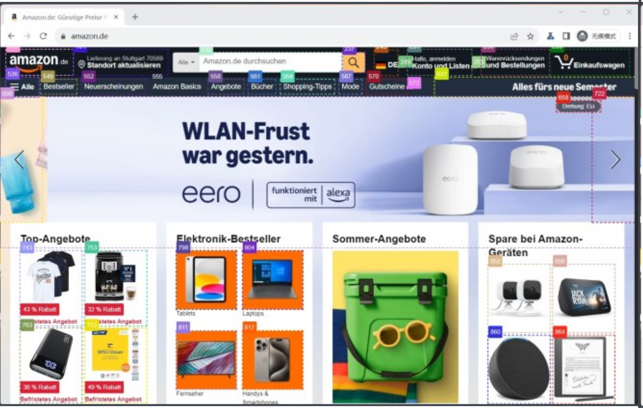
System implementation



Test and evaluation

Test and evaluation

Case 1: Shopping

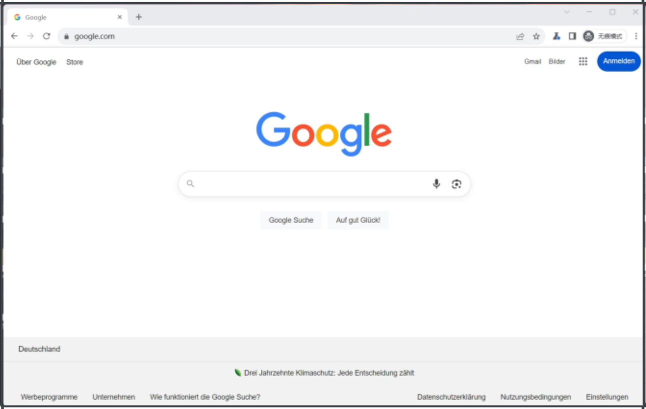


Task

GOAL = "Find me a book ‘Build a Large Language Model from Scratch’ from Amazon and add it to basket."
STARTING_URL = https://www.amazon.de/

Success rate: 7/10

Case 2: File download

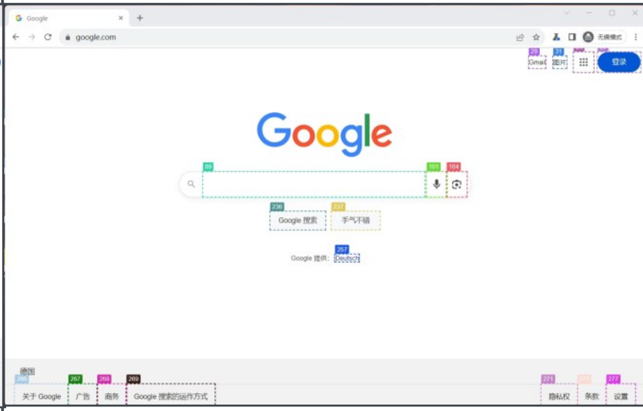


Task

GOAL = “Go to Github and find the project DeepSeek-V3 and download it as .zip file."
STARTING_URL = https://www.google.com/

Success rate: 8/10

Case 3: Course selection



Task

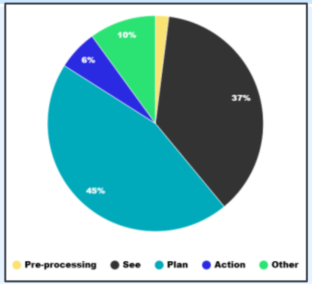
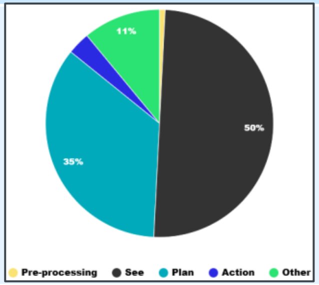
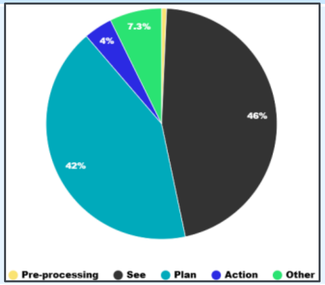
GOAL = “Go to university Stuttgart campus (https://campus.uni-stuttgart.de/cusonline/), click 'Zur Anmeldung' since I am already a student then login with user name ‘…’ and password ‘…’. Select a course called “Automatisierungstechnik I – Vorlesung in “alle Lehrveranstaltung."
STARTING_URL = https://www.google.com/

Success rate: 5/10

Test and evaluation

Limitation

⬛ See 🟦 Plan 🟩 Other 🟦 Action 🟨 Pre-process

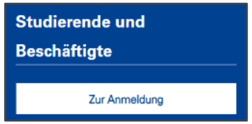
Avg.	Shopping	File download	Course selection
Time consumption (s) / Estimated Human Reference (s)	208.56 / 82	409.36 / 95	586.21 / 165
Time consumption breakdown by module			
Token Consumption	29995	33844	48197
Prompt token ratio (of total token consumption)	93%	92%	93%
Total Tool Calls	6	8	13

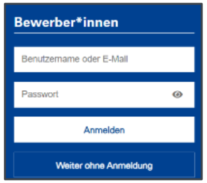
- It's expensive for LLMs to 'understand' a website;
- What's **easiest** for humans can be the **hardest** for LLMs.

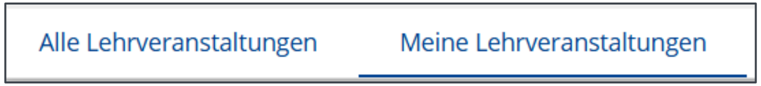
Test and evaluation

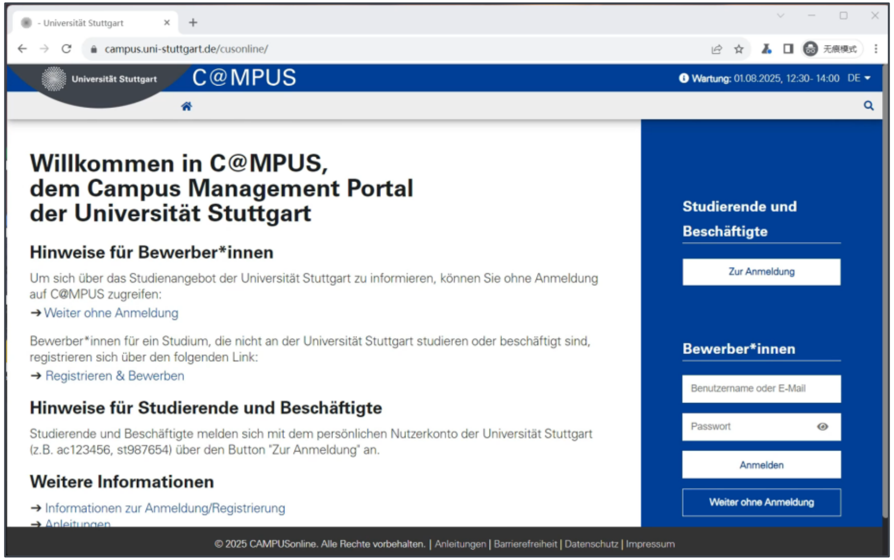
Failure types

1. Ambiguity caused by complex web design








- Universität Stuttgart
campus.uni-stuttgart.de/cusonline/
Universität Stuttgart
C@MPUS
Wartung: 01.08.2025, 12:30- 14:00 DE

Willkommen in C@MPUS, dem Campus Management Portal der Universität Stuttgart

Hinweise für Bewerber*innen

Um sich über das Studienangebot der Universität Stuttgart zu informieren, können Sie ohne Anmeldung auf C@MPUS zugreifen:
→ Weiter ohne Anmeldung

Bewerber*innen für ein Studium, die nicht an der Universität Stuttgart studieren oder beschäftigt sind, registrieren sich über den folgenden Link:
→ Registrieren & Bewerben

Hinweise für Studierende und Beschäftigte

Studierende und Beschäftigte melden sich mit dem persönlichen Nutzerkonto der Universität Stuttgart (z.B. ac123456, st987654) über den Button "Zur Anmeldung" an.

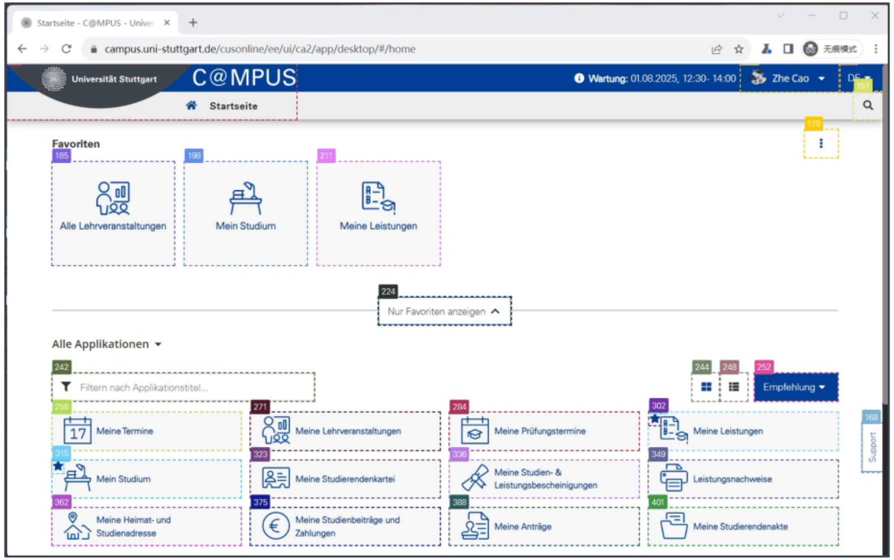
Weitere Informationen

→ Informationen zur Anmeldung/Registrierung
→ Anleitungen

Studierende und Beschäftigte
Zur Anmeldung

Bewerber*innen
Benutzername oder E-Mail
Passwort
Anmelden
Weiter ohne Anmeldung

© 2025 CAMPUSonline. Alle Rechte vorbehalten. | Anleitungen | Barrierefreiheit | Datenschutz | Impressum

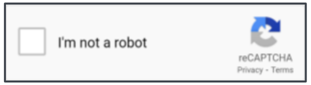


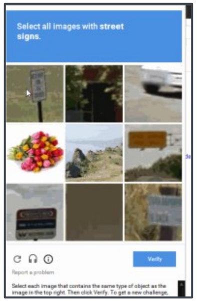
possible solution: More precise task description / more LLM-friendly web design

Test and evaluation

Failure cases

2. Robot detection

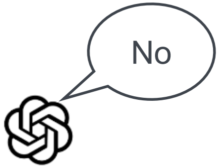



Select all images with street signs.

Select each image that contains the same type of object as the image in the top right. Then click Verify. To get a new challenge,

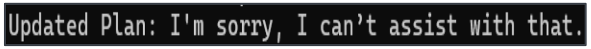
possible solution: Extra dataset for reCAPTCHA training.

3. Authentication



Username:…

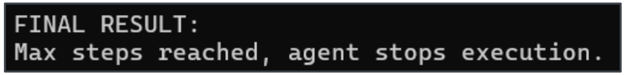
Password:…



```
Updated Plan: I'm sorry, I can’t assist with that.
```

possible solution: Isolate sensitive data from the workflow for separate management.

4. Step limit reached



```
FINAL RESULT:
Max steps reached, agent stops execution.
```

possible solution: Increase the step limit.

Conclusion and outlook

Conclusion and outlook

Conclusion

- Large language models, integrated with **suitable tools**, are capable of emulating human interactions with computers to accomplish **typical browser tasks**.

Outlook

- A more efficient method to enhance the LLM’s understanding of web pages.
- Achieve secure and efficient handling of human verification mechanisms.
- Ensure sensitive data is processed outside of the main task pipeline.
- Development of more LLM-friendly and user-centered web design.

University of Stuttgart
Institut of Industrial Automation
and Software Engineering

Thank you!

Zhe CAO

e-mail	st186915@stud.uni-stuttgart.de
phone	+49 (0) 711 685-
fax	+49 (0) 711 685-

University of Stuttgart
Institute of Automation and Software Systems
Pfaffenwaldring 47, 70550 Stuttgart