Patient-Specific Care Pathway Visualisation for Medical Nursing Staff

Sophie Bauernfeind, Selina Adlberger, Oliver Krauss, Andreas Stöckl
Research & Development, Research & Development, Digital Media Lab, Digital Media Lab,
University of Applied Sciences Upper Austria
Hagenberg, Austria
{sophie.bauernfeind, selina.adlberger, oliver.krauss, andreas.stoeck}@fh-hagenberg.at

Abstract

Nursing staff are increasingly confronted with extensive and detailed patient documentation, requiring much time to read through numerous possible care measures. Combined with rising patient loads, this underscores the need for a clearer and more immediately accessible overview of each patient's situation. Patient-specific care pathway visualisations offer a promising approach to reduce cognitive load, support faster decision-making, and improve situational awareness. This work investigates two Artificial intelligence (AI)-assisted methods for generating such visualisations: (1) simple image generation based on structured textual prompts, and (2) automated code generation that produces graph-based representations of clinical pathways. Using a dataset of synthetic patient profiles and seven defined care pathways, evaluating multiple state-of-the-art foundation models. The results highlight clear differences between models and approaches, particularly in language sensitivity, structural consistency, and the level of detail achievable. Image-based outputs provided visually rich overviews but frequently introduced subtle logical inconsistencies, while code-based methods produced verifiable and structurally coherent pathways yet varied in their ability to preserve contextual and psychosocial information. Together, these findings indicate that AI-assisted visualisation can effectively support—but not yet fully automate—patient-specific pathway generation, and they point toward hybrid solutions that combine visual accessibility with logical robustness.

Keywords: care pathways, clinical decision support, generative AI, visualization, elderly care

1. Introduction

In modern healthcare systems, nursing staff face increasing pressure as they manage growing patient loads, complex care requirements, and rapidly changing clinical environments (Jennings et al., 2022; Rachel and Francesco, 2018). A substantial portion of the nursing workload arises from the considerable time required to review extensive patient documentation and identify the most appropriate next steps in care. Nurses are frequently required to extract and synthesize information from numerous, often lengthy records to construct an accurate and up-to-date understanding of a patient's condition.

Visualisation and visualizing patient-specific care pathways has emerged as a promising strategy to address these challenges (Stadler et al., 2016; Mould et al., 2010; Tan et al., 2025). Instead of relying on static or generic process models, patient-specific visualizations dynamically reflect the individual's medical history, current interventions, and expected future steps. Creating these tailored visualizations remains a non-trivial task. Clinical pathways are complex, vary between patients, and evolve over time, which makes automated, flexible, and context-sensitive generation essential.

In this work, methods for generating patient-specific care pathway visualizations that can be seamlessly integrated into nursing workflows are investigated. Specifically, two methodological approaches got explored: (1) simple image generation, which allows creation of illustrations without requiring technical expertise, and (2) automated code generation, which produces graph representations using visualization libraries. The former prioritizes accessibility, enabling staff to obtain illustrative overviews with minimal effort, while the latter provides precision, adaptability, and integration potential with clinical data systems.

By comparing these two approaches, a better understanding is sought of how Artificial intelligence (AI)-assisted visualization generation can support nursing staff in managing complex care processes.

2. Related Work

Patient journey mapping provides structured visual descriptions of patients' movements through the healthcare system, helping clinicians identify barriers, transitions, and needs across the care continuum (Bulto et al., 2024). Such representations offer an important foundation for developing individualized visualizations that support rapid comprehension (Wagenaar et al., 2026).

Studies on pathway modeling emphasize the role of diagrammatic representations in clinical process improvement. Mould et al. demonstrate that accessible pathway diagrams can assist staff in identify-

Proceedings of the Third Workshop on Patient-Oriented Language Processing (CL4Health), pages 117–131
12 May 2026. ©ELRA Language Resources Association (ELRA), 2026

ing inefficiencies and evaluating redesign options, thereby improving communication and supporting organizational learning (Mould et al., 2010). Their findings highlight the practical relevance of clear visual models in everyday clinical workflows.

Gartner et al. propose an integrative framework that accounts for clinical, organizational, and informational dimensions, underscoring the need for pathway representations that reflect the complex realities of patient care and interprofessional collaboration (Gartner et al., 2022).

Manktelow et al. summarize methods for extracting time-resolved sequences of care activities from administrative and clinical data, showing how empirical pathway structures can support planning, quality improvement, and decision-making (Manktelow et al., 2022). These approaches demonstrate the potential for automated generation of patient-specific pathways grounded in actual care delivery.

Existing work on generative graph and visualization models has demonstrated strong performance, but these approaches typically rely on datasets that are already visualized or highly pre-structured. Automated data-to-visualization systems, for example, operate on clean tabular inputs that can be directly mapped into plotting code rather than inferred from narrative descriptions (Wu et al., 2024; Zou et al., 2024; Wang et al., 2025).

3. Methods

This study investigates two complementary approaches for generating patient-specific care pathway visualizations that support nursing staff in rapid situation assessment: (A) simple image generation and (B) automated code generation for graph-based visualizations. Both approaches are evaluated with respect to feasibility, interpretability, adaptability, and potential for integration into clinical workflows. To ensure a fair comparison, a shared data foundation and a common set of evaluation criteria are employed.

3.1. Care Pathway

A total of seven distinct clinical guidelines are utilized, each comprising between 5,771 and 126,403 words. These pathways encompass the following thematic areas: (1) Work Aid, (2) decubitus prophylaxis, (3) Dementia, (4) Urinary Incontinence, (5) Chronic Wound Care, (6) Pain Management, and (7) Fall Prevention.

3.2. Patient Data

In this study, a dataset consisting of ten synthetic patient profiles were generated with GPT-4. Each profile is provided as a plain text document structured around fourteen information categories.

These categories encompass: (0) general patient information such as name, gender, and age; (1) communication abilities, including hearing, vision, speech, and social interaction; (2) mobility, covering aspects such as movement in bed, and use of assistive devices; (3) vital functions including blood sugar management, medication handling, and pain assessment; (4) self-care activities, ranging from independent or assisted hygiene routines to skin and oral care; (5) eating and drinking behaviors, including levels of independence and support required for meal preparation and intake; (6) toileting abilities, continence status, and the use of catheters or related aids; (7) dressing and undressing capabilities, including the need for support and use of prostheses or orthoses; (8) sleep patterns, disturbances, and related observations; (9) hobbies and leisure activities; (10) preferences regarding the gender of caregiving staff as well as individual care goals and planned interventions; (11) safety-related aspects, such as orientation and the ability to recognize hazards; (12) household-related competencies, including heating, pet care, cleaning responsibilities, and relevant observations; and (13) personal expectations and wishes, religious considerations and preferences of both patients and their trusted persons, shown in Listing 1.

This dataset provides a consistent basis for comparing visualization approaches across varied patient situations, with documents ranging from 1,631 to 1,917 words.

3.3. Data Preparation and Pre-processing

Given the variability in patient data, not all treatment guidelines are applicable to every individual case. To streamline processing and ensure the accurate generation of visualization graphs, a two-stage pre-processing pipeline to structure both the clinical logic and the patient-specific relevance was implemented.

Guideline Structuring (Rule Extraction). Original clinical guidelines are typically available as narrative text, which can lead to ambiguity during automated processing. To mitigate this, Gemini 3 Pro was used to transform the raw guideline texts into a JavaScript Object Notation (JSON) format. Since the source guidelines were provided in German, the extracted logic retains the original German medical terminology to prevent semantic loss during translation. In this file, complex care instructions were decomposed into atomic "IF-THEN" statements:

- **Condition (IF):** Trigger criteria based on patient assessment (e.g., "IF Braden score is low").

```
First name: Elisabeth
Last name: Müller
Address: Musterstraße 10, 1010 Vienna
Date of birth: April 12, 1935
Main diagnoses: Dementia, high blood pressure, osteoarthritis
[...]
Care level: 4, application for increase submitted: no
Emergency call system: yes, with Johanniter Unfallhilfe
[...]
Speech disorder: No
Type of speech disorder: -
Skills/resources/
Observations: Despite her dementia, Elisabeth is still able to hold simple
↪  conversations. She responds positively to friendly and patient communication.
↪  She tends to repeat herself and has difficulty remembering new situations and
↪  faces. She shows confusion when discussing complex topics. She likes being read
↪  to and finds it relaxing.

2. ABEDL Moving around Independently    Supervision    Instruction    Partial
↪  assistance    Full assistance    Relevant for care plan.

Movement in bed (changing position/sitting up):                  X
[...]
Bowel movements: regular    irregular   X normal      diarrhea      constipation
[...]
```

Listing 1: Excerpt of a patient profile, translated to provide reader clarity. Full original record contains 1,773 words.

- **Action (THEN):** The mandatory nursing intervention (e.g., "THEN initiate positioning schedule").

This transformation reduces the input tokens for the language model by converting text into a clearer and cleaner structure. An excerpt of the generated structure is shown in Listing 2.

Patient-Guideline Mapping. Since applying all guidelines to every patient would create cluttered and irrelevant visualizations, an intermediate mapping step was introduced. Gemini 3 Pro was used to automatically generate a specific mapping file that acts as a relevance filter. The model analyzed the unstructured patient profile to identify relevant diagnoses and care needs, subsequently flagging specific guideline modules as `active: true` or `active: false` (e.g., activating "Dementia Care" only for patients with a confirmed diagnosis). This ensures that the generated pathways focus exclusively on clinically indicated interventions.

3.4. Experimental Models

To evaluate the two visualization approaches, state-of-the-art foundation models were employed. Models were allocated to Approach A (Image Generation) and Approach B (Code Generation) as shown in Table 1. All models were tested with prompts in both English and German to evaluate cross-lingual robustness.

Model	A1	A2	B1	B2
Gemini 3 Pro	X	X	X	X
Gemini 3 Fast	–	–	X	X
Gemini 3 Thinking	X	X	–	–
NanoBanana Pro	X	X	–	–
GPT-5	–	–	X	–

Table 1: Model usage across test setup.

Experimental Design Framework. To systematically assess both visualization approaches, an evaluation framework was developed. Within this framework, three key parameters were varied:

- **Data Source:** Pre-processed Logic (JSON) vs. Unstructured Full-Text (All Guidelines).
- **Model:** Depending on the approach, models are used as listed above.
- **Prompt Language:** German (Native Source) vs. English (Translation).

3.5. Approach A: Simple Image Generation

The first approach explores the capability of multimodal models to synthesize complex clinical data directly into visual artifacts, bypassing traditional rendering software. To ensure clinical utility, strict, role-based prompts were designed to enforce specific aesthetic and structural constraints (see Listing 3 and Listing 4).

```
{
"arbeitshilfe_complete_txt": [
  "IF Aufnahme eines neuen Patienten THEN Durchfhrung der SIS (Strukturierte
   ↪  Informationssammlung) innerhalb von 24h.",
  "IF Risikoeinschtzung (Matrix) zeigt Handlungsbedarf THEN Planung konkreter
   ↪  Manahmen im Manahmenplan.",
  "IF Manahmen durchgefhrt THEN Dokumentation im Berichteblatt (Abweichungen
   ↪  begrnden).",
  "IF Evaluationsdatum erreicht OR akute Zustandsnderung THEN Erneute Einschtzung
   ↪  der Risiken und Anpassung des Manahmenplans.",
  "IF Pflegegrad 4 oder 5 THEN Tgliche Dokumentation der Krperpflege und Mobilitt
   ↪  zwingend erforderlich."
]
}
```

Listing 2: Excerpt of the pre-processed guideline logic in german. Explicit IF-THEN rules were extracted by Gemini 3 Pro to standardize the input for downstream visualization.

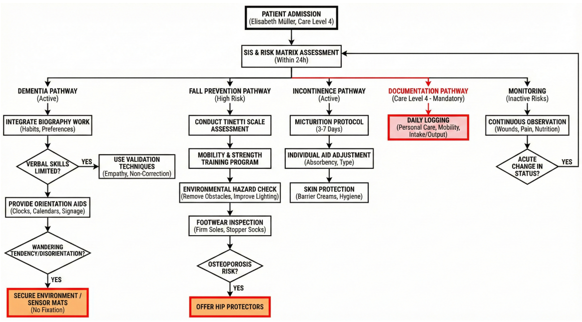


Figure 1: Generated image from NanoBanana Pro, with English prompt.

3.5.1. Input Variation A1: Structured Guideline Injection

In this setup,the model used pre-processed "If-Then" rules (`guidelines.json`) and a patient filter (`guidelines_patients.json`) to generate a decision tree (see Listing 3). The objective was to test whether providing structured logic improves the accuracy and consistency of the visual pathways.

3.5.2. Input Variation A2: Unstructured Full-Text Injection

In this variation, the model received the complete, unprocessed textual guidelines without specific visual constraints (see Listing 4). The objective was to assess its ability to autonomously extract clinical rules from raw text, resolve ambiguities, and generate coherent pathways.

Image Synthesis. The visualization process was orchestrated by Gemini 3 Pro, utilizing the NanoBanana Pro model as the generative engine within the Gemini environment (see Figure 1). Gemini 3 Pro first processed the prompt and patient data to structure the visual request, subsequently instructing NanoBanana Pro to generate the output. The model was instructed to emphasize a decision tree's logical branching, placing text in boxes when possible while preserving the correct care-path topology.

```
System Role: You are a technical
 ↪  illustrator for medical algorithms.
 ↪  Create a highly detailed
 ↪  visualization of a clinical pathway
 ↪  as a flowchart.
HERE IS THE DATA:
### 1. PATIENT (The focus):[Insert
 ↪  contents of 'patient_0.txt']
### 2. ACTIVE TOPICS (The
 ↪  filter):[Insert contents of
 ↪  'guidelines_patients.json']
### 3. RULES (The logic):[Insert
 ↪  contents of 'guidelines.json']
IMAGE INSTRUCTIONS:
- Generate a strict "flowchart" on a
 ↪  white, technical background.
- Show clearly defined nodes (boxes) and
 ↪  connecting lines (arrows).
- Structure: At the top is the starting
 ↪  point "patient admission," below
 ↪  which the tree branches into
 ↪  different paths (fall, dementia,
 ↪  care).
- Style: Vector graphics, schematic,
 ↪  clean, medically precise. No photos,
 ↪  no 3D effects.
- Colors: Black/white with signal colors
 ↪  (red/orange) for risk paths.
Additional note for AI: Try to suggest
 ↪  text in the boxes, but focus
 ↪  primarily on the correct logical
 ↪  branching aesthetics of a decision
 ↪  tree.
```

Listing 3: Prompt template for Approach A1. The model receives structured JSON rules and specific color signaling instructions to generate the visualization.

```
System Role: You are a technical
↪  illustrator specializing in medical
↪  algorithms. Create a detailed
↪  visualization of a clinical pathway
↪  in the form of a flowchart.
HERE IS THE DATA:
### 1. PATIENT (The Focus):[Insert
↪  contents of 'patient_0.txt']
### 2. ACTIVE ISSUES:[Insert contents of
↪  'guidelines_patients.txt']
### 3. RULES (The Logic):[Insert
↪  unstructured raw guideline texts]
IMAGE INSTRUCTIONS:
- Create a strict "flowchart" on a
↪  white, technical background.
- Show clearly defined nodes (boxes) and
↪  connecting lines (arrows).
- Structure: At the top is the starting
↪  point "Patient Admission"; below
↪  that, the tree branches into
↪  different paths (Fall, Dementia,
↪  Care).
- Style: Vector graphics, schematic,
↪  clear, medically precise. No photos,
↪  no 3D effects.
```

Listing 4: Prompt template for Approach A2. The model must extract visual logic directly from unstructured text guidelines without specific color signaling constraints.

Quality Assurance Criteria. Generated outputs were screened against:

1. **Risk Visibility:** Whether patient-specific risk factors (e.g., fall risk, dementia-related supervision needs) were clearly represented and correctly highlighted according to the defined visual signaling scheme.
2. **Data Faithfulness:** Whether the visual branching logic corresponded to the injected patient data and guideline content, without omission of central care elements.
3. **Occurrence of Hallucinated Elements:** Whether the visualization contained invented nodes, incorrect causal links, dangling edges, or visual artifacts not grounded in the provided input data.

3.6. Approach B: Code-Generated Graph Visualizations

The second approach prioritizes interoperability and lightweight system integration by generating Mermaid.js code (Sveidqvist and the Mermaid.js Contributors, 2014). Unlike complex rendering scripts, Mermaid offers a declarative, markdown-compatible syntax that allows care pathways to be embedded directly into clinical dashboards as renderable text.

3.6.1. Input Variation B1: Structured Logic

In this setup, Gemini 3 Pro, Gemini 3 Fast, and GPT-5 received the `guidelines.json` containing explicit "If-Then" rules alongside the patient profile (see Listing 5). The objective was to test the models' ability to strictly follow syntax constraints and generate valid Mermaid flowchart syntax when the logic extraction is already pre-processed.

```
System Role: You are a medical
↪  assistant. Create a precise process
↪  graph (Mermaid.js) for care
↪  planning.
Use these three data sources:
### SOURCE 1: THE PATIENT (status &
↪  vitals): [Insert contents of
↪  'patient_0__full.txt']
### SOURCE 2: RELEVANT TOPICS (Filter):
↪  [Insert contents of
↪  'guidelines_patients.json']
### SOURCE 3: LOGIC & RULES (Knowledge
↪  Base): [Insert contents of
↪  'guidelines.json']
YOUR TASK: Create a flowchart that
↪  logically deduces what needs to be
↪  done.
Procedure:
1. Look in SOURCE 2: Which topics are
↪  active? (e.g., falls, dementia).
2. Go to SOURCE 3 (rules) for these
↪  topics.
3. Use the data from SOURCE 1 (patient)
↪  to check which "if" condition is
↪  met.
   - Example: Rule says "If Braden <
   ↪  12." According to Source 1, the
   ↪  patient has "Braden 10." ->
   ↪  Include the measure "soft
   ↪  bedding" in the graph.
   - If a condition is NOT met, omit the
   ↪  measure.
Just give me back the Mermaid code.
```

Listing 5: Prompt template for Approach B1. The model receives pre-extracted JSON rules and a concrete deduction example.

3.6.2. Input Variation B2: Unstructured Full-Text

In this setup, only the Gemini 3 Series models were tested. GPT-5 was excluded from this specific variation due to platform-imposed file upload limits that prevented the ingestion of the complete guideline set. As shown in Listing 6, the models had to perform implicit logic extraction directly from raw text. The objective was to evaluate the reasoning capabilities of the *Pro* versus *Fast* variants when filtering noisy, real-world data and structuring it into a valid Mermaid graph within a single context window.

```
System Role: You are a medical
↪  assistant. Create a precise process
↪  graph (Mermaid.js) for care
↪  planning.
Use these data sources:
### SOURCE 1: THE PATIENT (status &
↪  vitals): [Insert contents of
↪  'patient_0__full.txt']
### SOURCE 2: RELEVANT TOPICS (Filter):
↪  [Insert contents of
↪  'guidelines_patients.json']
### SOURCE 3: LOGIC & RULES (Knowledge
↪  Base): Use the necessary guidelines.
↪  [Insert unstructured raw guideline
↪  texts]
YOUR TASK: Create a flowchart that
↪  logically derives what needs to be
↪  done.
Procedure:
1. Look at SOURCE 2: Which topics are
↪  active? (e.g., falls, dementia).
2. Go to the appropriate guideline for
↪  these topics.
3. Use the data from SOURCE 1 (patient)
↪  to check which condition is met.
   - If a condition is NOT met, omit the
   ↪  measure.
Just give me back the Mermaid code.
```

Listing 6: Prompt template for Approach B2. The model must extract logic directly from raw text guidelines without predefined examples.

Evaluation Criteria. The generated Mermaid code was assessed based on:

1. **Syntax Stability:** Whether the generated Mermaid code rendered successfully without manual correction and adhered to valid structural conventions.

2. **Data Fidelity:** Whether patient-specific risks, diagnoses, and contextual constraints were accurately reflected in the generated graph structure.

3. **Logical Coherence:** Whether the generated pathways followed a clinically meaningful sequence without contradictory or structurally inconsistent connections.

4. **Usability:** Whether the resulting graph provided a clear and interpretable overview suitable for rapid situation assessment in nursing workflows.

3.7. Validation Procedure

Both approaches were validated using the dataset of ten synthetic patient profiles. For each scenario, following steps were done:

1. **Execution Matrix:** Prompts were run for Approach A (Image) and Approach B (Code) across all selected models and both languages (English/German).

2. **Syntax Verification (Approach B):** For the code-based approach, it was assessed whether the generated Mermaid.js code rendered successfully without manual correction.

3. **Visual Inspection (Approach A):** For the multimodal approach, the generated dashboards were qualitatively assessed for hallucinations (e.g., legible text vs. gibberish, correct risk color-coding).

4. **Cross-Lingual Check:** The output was compared between German and English prompts to determine if translation layers introduce medical inaccuracies.

3.8. Synthesis and Decision Guidance

Findings are synthesized to identify:

- **Model Efficiency:** Trade-offs between latency-optimized models (e.g., Gemini 3 Fast) and high-reasoning variants (e.g., Gemini 3 Pro/Thinking) for reliable graph generation.

- **Data Robustness:** The comparative reliability of pre-processed logic (JSON) versus flexible full-text extraction.

- **Platform Viability:** The impact of commercial tool constraints (e.g., file limits in Copilot) on clinical workflow integration.

- **Format Suitability:** Guidelines for deploying direct visual generation versus structured Mermaid graphs based on clinical documentation needs.

4. Results

The performance of the generated visualizations was evaluated across ten synthetic patient profiles. The evaluation was carried out by two researchers with completed graduate-level study in the medical and bioinformatics fields. The results are categorized by the generation approach (Multimodal vs. Code-based) and analyzed regarding model capabilities and language robustness.

4.1. Approach A: Visual Fidelity and Semantic Consistency

The direct image generation using NanoBanana Pro produced visually distinct outputs depending on the prompt language and input granularity. While the requested aesthetic style was generally adhered to by the model, substantial variances in information density and correctness were observed.

Language-Dependent Stability. A comparative analysis of German versus English prompts (using identical JSON inputs) revealed a marked disparity in output quality:

- **German Prompts:** The outputs exhibited suboptimal spatial organization and high visual noise. The generated dashboards frequently suffered from orthographic instability (hallucinated or misspelled text labels) and displayed a lack of content specificity, often resorting to generic medical imagery rather than patient-specific data points.

- **English Prompts:** These yielded superior structural coherence and orthographic fidelity, with no detected spelling errors. Furthermore, the semantic mapping of risk factors was more precise, evidenced by the correct application of the requested signal color scheme (red/orange) for high-priority pathways.

Complexity vs. Logical Grounding (Full-Text Input). When utilizing Gemini 3 Pro to drive the image generation process based on unstructured full-text guidelines (variation A2 3.5.2), the resulting visualizations increased in graphical complexity. Although the visual output was rendered by NanoBanana Pro, the guiding logic provided by Gemini 3 Pro led to more detailed structures.

- **Superficial Plausibility:** At a macroscopic level, the charts appeared highly detailed and aesthetically professional ("superficial coherence").

- **Semantic Disconnects:** However, a granular inspection revealed severe topological errors, such as dangling edges and erroneous causal links (see Figure 2), indicating a hallucination of logic despite the high visual resolution.

4.2. Approach B: Code-Generated Graph Visualizations

The generated Mermaid.js code was evaluated with respect to visual usability, procedural logic, and information fidelity. The analysis reveals that model architecture and prompt language considerably dictate the utility of the resulting graph, creating a trade-off between overview speed and safety-critical detail.

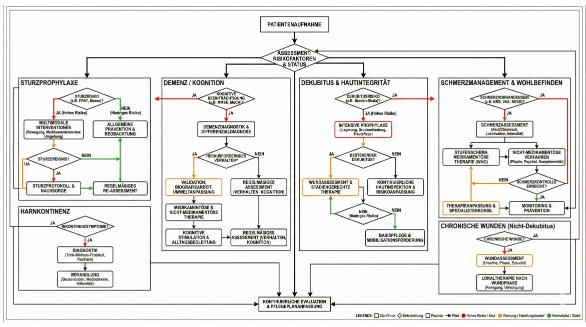


Figure 2: Generated image using Gemini 3 Pro and NanoBanana Pro with German prompt. The output illustrates severe "visual noise" and topological hallucinations, with non-linear pathways and disconnected logical vectors.

4.2.1. Visual Layout and Abstraction

Gemini 3 Fast (German, Variant B1 3.6.1) demonstrated the highest utility for rapid situation assessment. It spontaneously generated styling classes, grouping related interventions into color-coded blocks and omitting inactive pathways. This resulted in a compact "dashboard view" ideal for the point of care.

In contrast, English prompts resulted in unstyled graph topologies with varying degrees of clutter. Gemini 3 Fast (English, Variant B1 3.6.1) exhibited the lowest readability, generating dense cyclic structures with excessive feedback loops (e.g. evaluation arrows returning to the start), creating visual noise. Gemini 3 Pro (English, Variant B1 3.6.1) adopted a structured "Assessment $\rightarrow$ Action" pattern. While visually clearer than the Fast model, it still listed non-applicable guidelines as isolated nodes. Although it avoided expanding their internal logic, these inactive headers occupied unnecessary visual space, unlike the German output which completely omitted inactive pathways.

However, a notable improvement was observed in Gemini 3 Pro (English, Variant B2 3.6.2). This configuration produced the most aesthetically coherent output of all evaluated permutations, offering a clean, streamlined flow. The primary drawback of this visualization was minor rendering artifacts, where text labels occasionally overlapped with directional arrows, slightly obscuring the path logic (see Figure 3).

4.2.2. Procedural Logic and Precision

Gemini 3 Pro (German, Variant B1 3.6.1) demonstrated the highest procedural fidelity. It adopted a precise "Question $\rightarrow$ Action" pattern and successfully integrated social resources directly into the clinical workflow, making it the most suitable model for generating concrete nursing instructions.

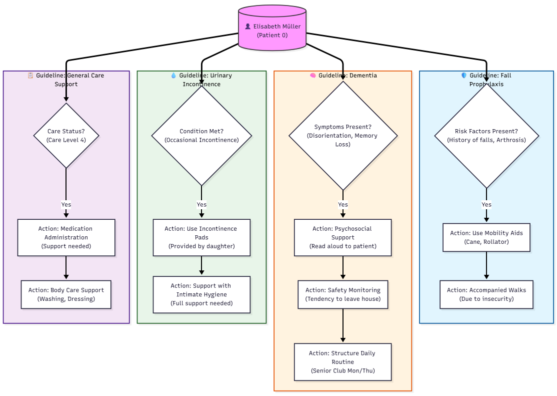


Figure 3: Generated image using Gemini 3 Pro with English prompt and full unstructured text input, showing improved layout but minor text overlap.

In contrast, GPT-5 (Variant B1 3.6.1) produced rigid, list-based structures. Despite the structured input, it listed all theoretically possible measures regardless of specific patient needs, resembling a static checklist rather than a dynamic pathway.

4.2.3. Information Retention and Cross-Lingual Nuances

The qualitative analysis of graph content revealed a substantial interaction between model architecture and prompt language regarding the preservation of psychosocial "soft factors."

- **Narrative Retention vs. Abstraction:** A disparity in the level of detail was observed between the efficiency-focused and high-performance models. Gemini 3 Fast (German, Variants B1 3.6.1 and B2 3.6.2) and GPT-5 demonstrated a tendency toward "terminological abstraction", generalizing specific patient resources (e.g., the calming ritual of reading aloud) into generic nursing categories. In contrast, Gemini 3 Fast (English) maintained higher narrative fidelity, explicitly incorporating the specific action.

- **Model Robustness:** Gemini 3 Pro exhibited the highest degree of information stability across both languages. It successfully mapped the specific biographical resources to the corresponding intervention nodes without loss of detail, regardless of whether the prompt was in German or English. This suggests that higher-parameter models effectively mitigate the abstraction risks inherent in smaller models.

- **Linguistic Stability and Code-Switching:** Despite the high narrative retention in English, Gemini 3 Fast (English, Variants B1 3.6.1 and B2 3.6.2)) showed signs of "translation leakage". While the graph logic was correctly translated, specific technical terms for medical aids remained untranslated within the English node structure ("Promote use of Rollator/Gehstock"). This phenomenon of code-switching indicates a shallower translation layer in efficiency-optimized models compared to the Pro-tier architecture.

4.3. Platform Constraints

A decisive factor in the evaluation was platform accessibility. While the Gemini environment allowed for the seamless processing of the complete guideline set in variation B2 (3.6.2), GPT-5 failed to process the full-text scenario due to strict file upload limits, necessitating the exclusive use of the preprocessed JSON approach for this model.

5. Discussion

This investigation into AI-assisted care pathway visualization highlights trade-offs between aesthetic fidelity, structural verifiability, and preservation of patient-specific context. Analysis of generation errors revealed modality-specific vulnerabilities. In Approach A (image generation), reliance on pixel-based models led to superficial plausibility, with visually coherent outputs often containing topological inconsistencies such as incorrect connections and unsupported clinical elements. In contrast, Approach B (code generation) produced structurally verifiable outputs, as errors typically resulted in rendering failure rather than misleading visuals. However, this approach introduced semantic limitations, as smaller models tended toward terminological abstraction, generalizing individualized interventions and thereby losing important person-centered details. While generative models show promise in reducing nursing cognitive load, direct end-to-end generation remains insufficient for safety-critical applications without robust human oversight.

5.1. The Trap of Superficial Plausibility

A primary finding regarding Approach A (Direct Image Generation) is the discrepancy between visual polish and logical grounding. Models like NanoBanana Pro, even when guided by high-reasoning prompts (Gemini 3 Pro), produced charts with high "superficial plausibility" they visually resembled professional clinical algorithms. However, the presence of "hallucinated logic", such as dangling edges or causally incorrect connections, poses a non-trivial risk. Unlike code-based generation (Approach B), where errors typically manifest as syntax failures (rendering the output blank and

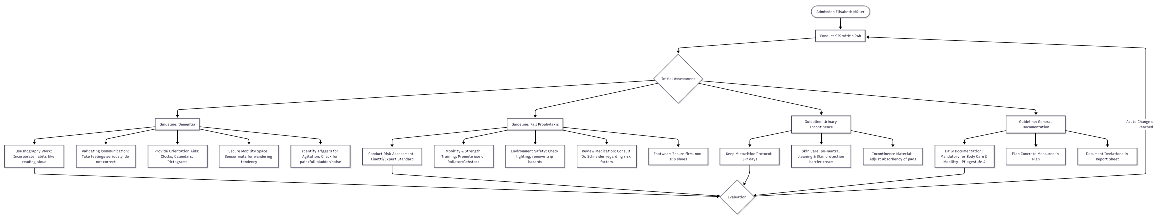


Figure 4: Generated graph code using Gemini 3 Fast with English prompt and structured JSON input.

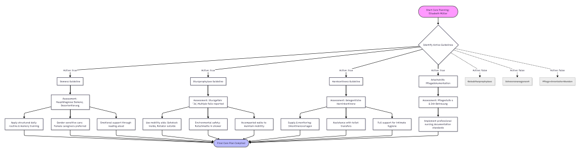


Figure 5: Generated graph code from Gemini 3 Fast with english promt and unstructured full text input, demonstrating untranslated terminology from the source text.

thus harmless), image generation errors are subtle and integrated into the visual artifact.

5.2. Language-Driven Abstraction vs. Narrative Fidelity

The comparison of prompt languages (English vs. German) uncovered a nuanced trade-off between structural stability and the granularity of care instructions.

- **Syntactic Stability and Code-Switching:** English prompts consistently yielded more stable Mermaid.js code, confirming that models are better attuned to logical control structures in English. However, efficiency-focused models like Gemini 3 Fast exhibited "translation leakage," where technical German terms remained untranslated within the English syntax.

- **Narrative Retention vs. Professional Abstraction:** The results indicate that the native language (German) often triggers a "professional filter" in smaller models. Models like Gemini 3 Fast and GPT-5 tended to abstract specific rituals into generic professional terminology.

This suggests that for clinical applications where the preservation of psychosocial nuances is as critical as medical accuracy, the use of high-performance models (Pro-tier) is mandatory. Efficiency models, while faster, introduce a "granularity risk" where essential person-centered details are sacrificed for either professional abstraction (in native language) or linguistic instability (in English).

6. Conclusion

The evaluation of two AI-assisted approaches for generating patient-specific care pathway visualizations shows that both offer benefits but also notable limitations. Image-based generation provides visually appealing outputs but lacks reliable logical grounding, making it unsuitable for safety-critical contexts. Code-based generation yields structurally verifiable pathways, though results vary with model architecture and prompt language. English prompts favor syntactic stability, whereas German prompts better preserve contextual and psychosocial details.

Overall, while AI-driven visualization can support nursing workflows, neither approach is yet sufficient as an autonomous solution. Future systems should integrate the strengths of both methods, combining visual clarity with semantic and contextual accuracy to more effectively reduce cognitive burden and enhance care efficiency.

7. Outlook

In this setup we tested the use of generated images or graph visualizations without a bigger framework. The results suggest that a useful tool that is easy to use requires more structure and refinement. The tool's output should be as consistent as possible so that nursing staff can familiarize themselves with the flowchart and thus access the information more quickly. This would require a framework which combines machine learning and algorithmic processes.

8. Ethical Considerations

These experiments were conducted using synthetic data, which cannot fully reflect real-world conditions. The synthetic patient data was generated for research purposes only and does not represent real individuals. When real-world data is used, multiple safety measures have to be included. In Europe, for example, such minimum requirements are established and regulated by the European AI Act, the General Data Protection Regulation, and the Medical Devices Regulation. The aim is to protect individuals' data in the context of patient care. In order to develop a tool to support nursing staff, these regulations must be followed.

9. Limitations

The study also has several limitations:

1. **Synthetic data:** While the generator produces realistic patterns, real clinical records contain more variability and complexity.
2. **Comparability:** The image content created was evaluated visually by hand. No comparative metrics were used.
3. **Sample size:** The size of the patient sets is limited to 10 patients with 7 different care pathways, restricted by the manual effort required for comparability.
4. **No clinical validation:** This study does not include validation and clinical utility with healthcare providers.
5. **Model usage:** We used common known n pixel-based image generation models, other models based on logical relationship-based generation might deliver better results. Additionally the used models are paid models. To keep costs low and ensure data security, self-hosted models would also be worth considering for future evaluation.

10. Bibliographical References

Lemma N. Bulto, Ellen Davies, Janet Kelly, and Jeroen M. Hendriks. 2024. Patient journey mapping: emerging methods for understanding and improving patient experiences of health systems and services. *European Journal of Cardiovascular Nursing*, 23(4):429–433.

Jean-Baptiste Gartner, Kassim Said Abasse, Frédéric Bergeron, Paolo Landa, Célia Lemaire, and André Côté. 2022. Definition and conceptualization of the patient-centered care pathway: a proposed integrative framework for consensus. *BMC Health Services Research*, 22:558.

Bonnie Mowinski Jennings, Marianne Baernholdt, and Susan G. Hopkinson. 2022. Exploring the turbulent nature of nurses' workflow. *Nursing Outlook*, 70(3):440–450.

Matthew Manktelow, Aleeha Iftikhar, Magda Bucholc, Michael McCann, and Maurice O'Kane. 2022. Clinical and operational insights from data-driven care pathway mapping: a systematic review. *BMC Medical Informatics and Decision Making*, 22(1):43.

Gillian Mould, John Bowers, and M. Ghattas. 2010. The evolution of the pathway and its role in improving patient care. *BMJ Quality & Safety*, 19(5):e14.

Harrad Rachel and Sulla Francesco. 2018. Factors associated with and impact of burnout in nursing and residential home care workers for the elderly. *Acta Bio Medica : Atenei Parmensis*, 89(Suppl 7):60.

Jennifer G. Stadler, Kipp Donlon, Jordan D. Siewert, Tessa Franken, and Nathaniel E. Lewis. 2016. Improving the Efficiency and Ease of Healthcare Analysis Through Use of Data Visualization Dashboards. *Big Data*, 4(2):129–135.

Knut Sveidqvist and the Mermaid.js Contributors. 2014. Mermaid.js: Open-Source Diagramming and Charting Tool. `https://mermaid.js.org/`. Accessed: 2026-02-17.

Xunan Tan, Xiang Suo, Wenjun Li, Lei Bi, and Fangshu Yao. 2025. Data visualization in healthcare and medicine: a survey. *The Visual Computer*, 41(5):3037–3058.

Britt Wagenaar, Hanneke Merten, and Cordula Wagner. 2026. Positive and negative experiences and emotions of older patients before, during and after hospital admission: patient journey mapping. *BMC Geriatrics*.

Yu Wang, Ryan A. Rossi, Namyong Park, Nesreen K. Ahmed, Danai Koutra, Franck Dernoncourt, and Tyler Derr. 2025. Demystifying the Power of Large Language Models in Graph Generation. *Findings of the Association for Computational Linguistics: NAACL 2025*, page 8174–8189.

Yang Wu, Yao Wan, Hongyu Zhang, Yulei Sui, Wucai Wei, Wei Zhao, Guandong Xu, and Hai Jin. 2024. Automated data visualization from natural language via large language models: An exploratory study. *Proceedings of the ACM on Management of Data*, 2(3):1–28.

Bocheng Zou, Mu Cai, Jianrui Zhang, and Yong Jae Lee. 2024. VGBench: Evaluating Large Language Models on Vector Graphics Understanding and Generation. *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, page 3647–3659.

A. Generated Graph Data

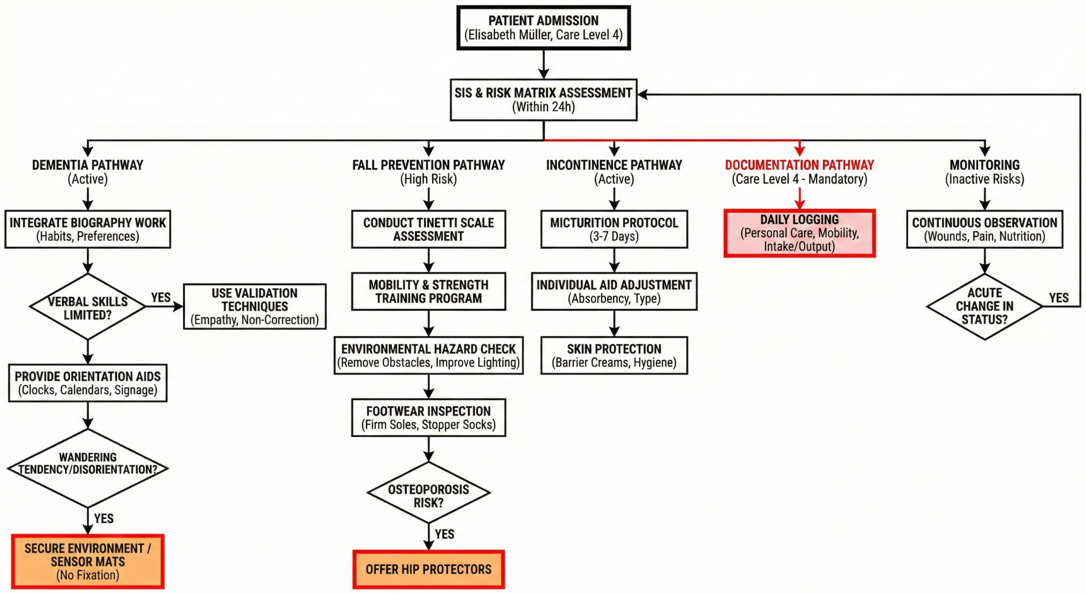


Figure 6: Bigger display of Figure 1, generated image using NanoBanana Pro, with English prompt.

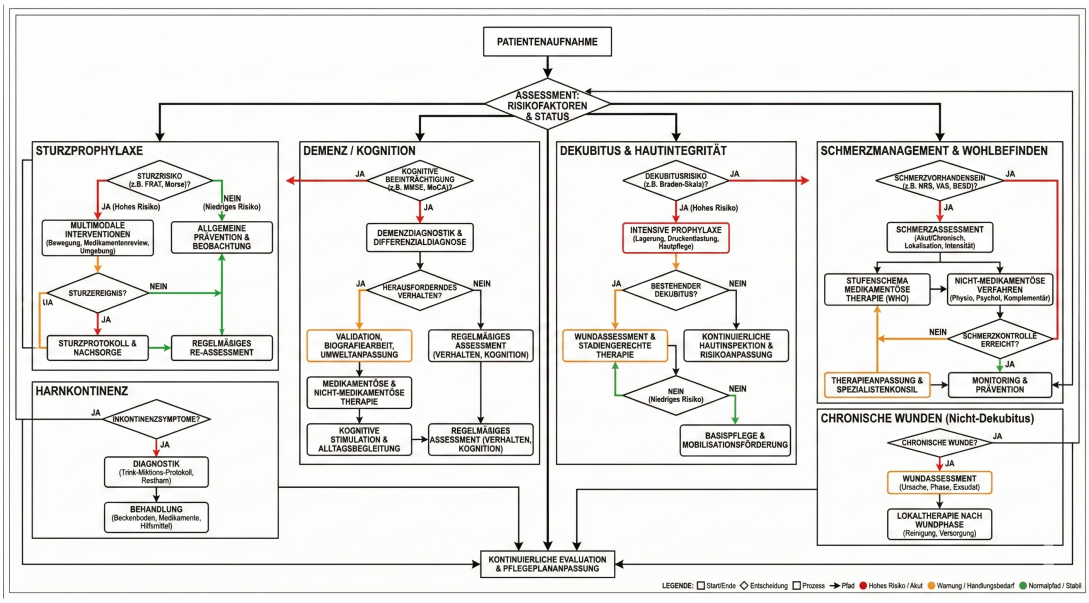


Figure 7: Bigger display of Figure 2, generated image using Gemini 3 Pro and NanoBanana Pro with German prompt.

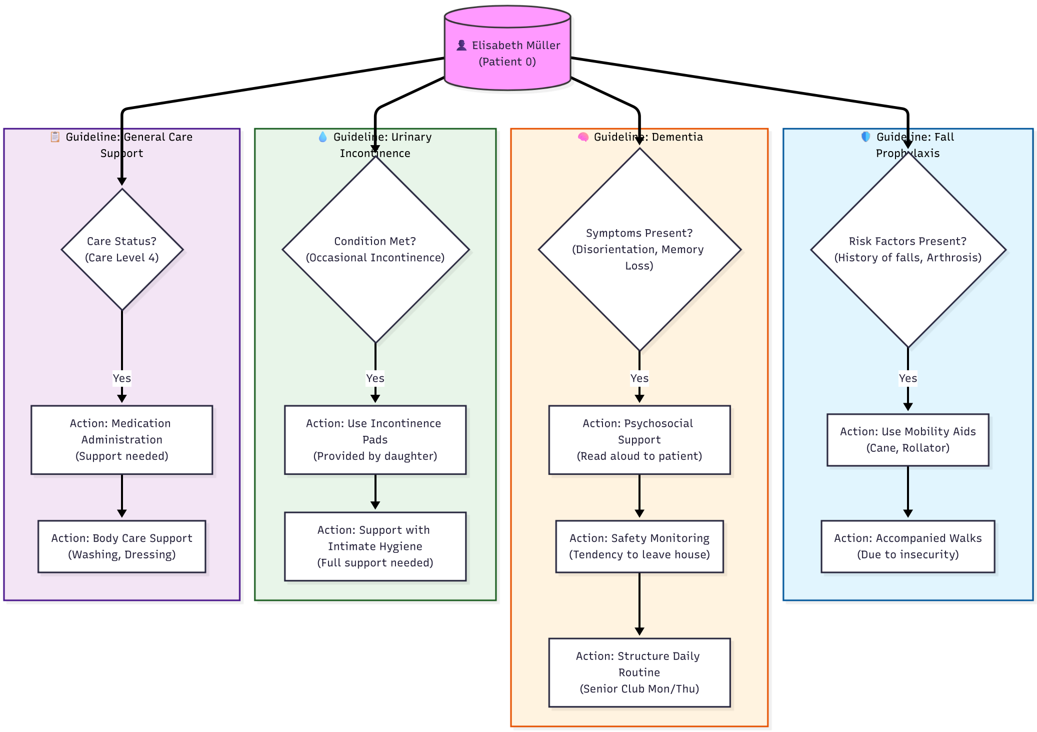


Figure 8: Bigger display of Figure 3, generated image using Gemini 3 Pro with English prompt and full unstructured text input, showing improved layout but minor text overlap.

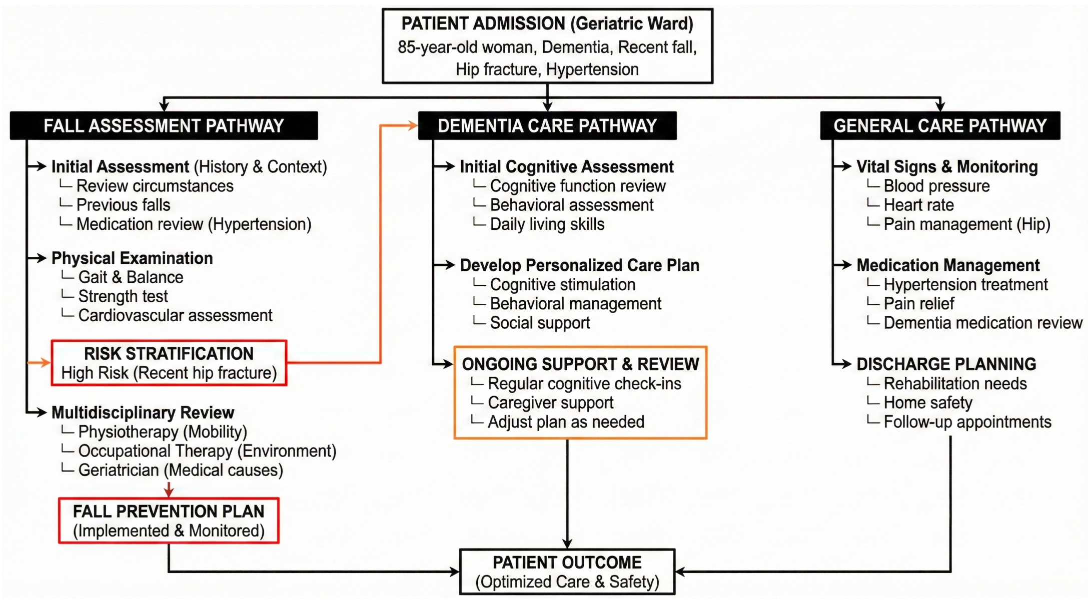
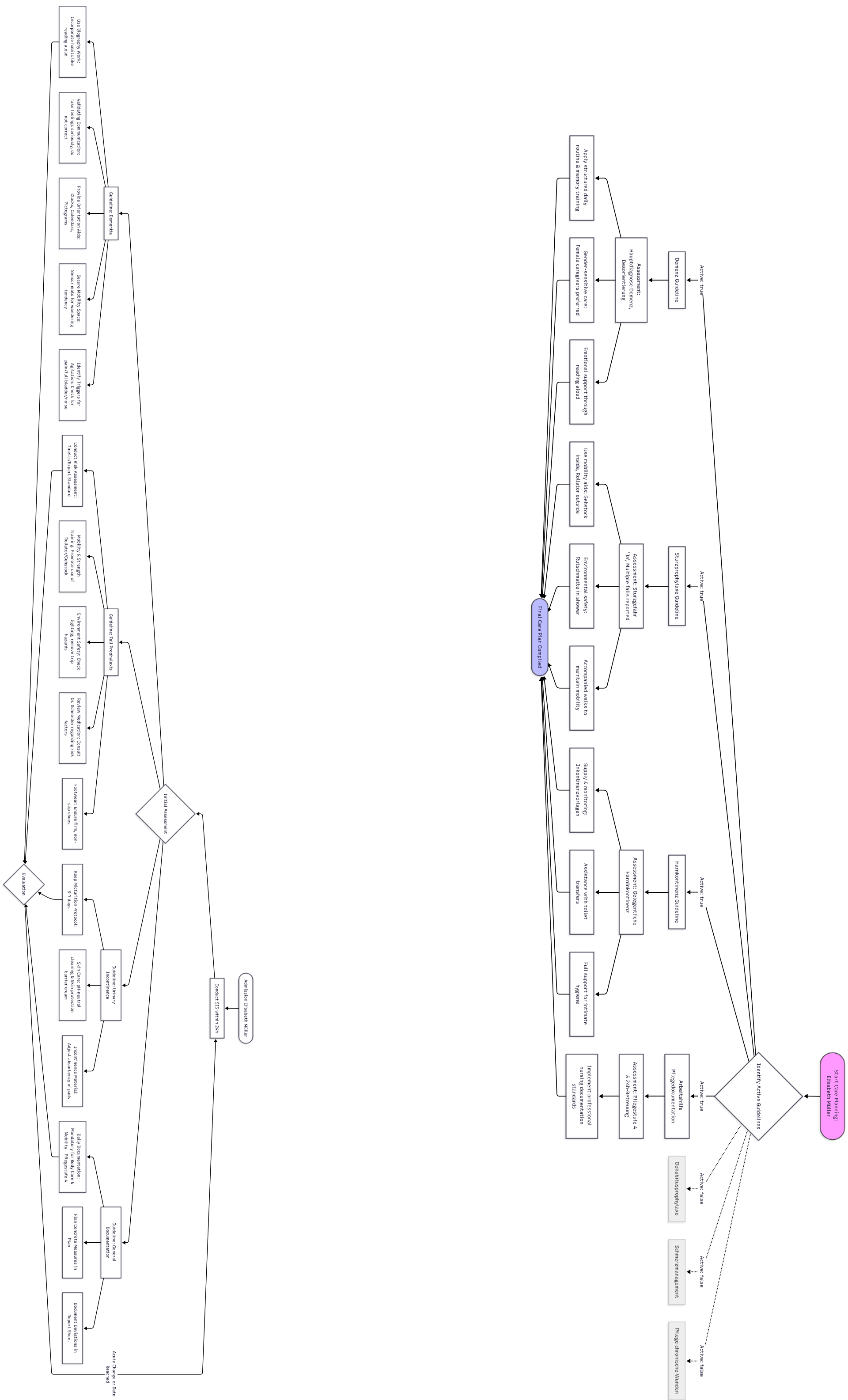


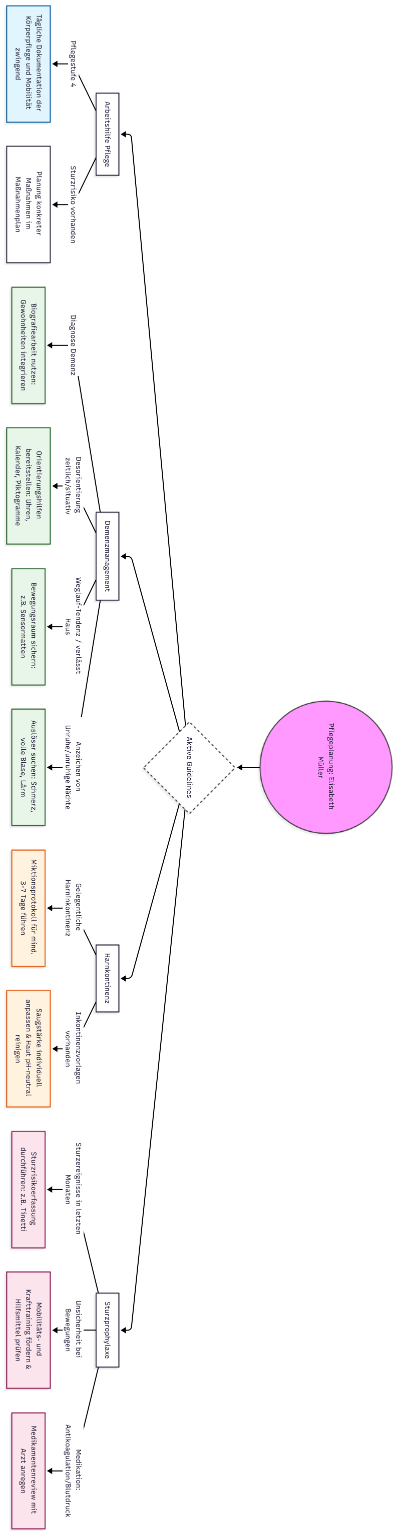
Figure 9: Generated image using Gemini 3 Pro with English prompt and structured guideline input.



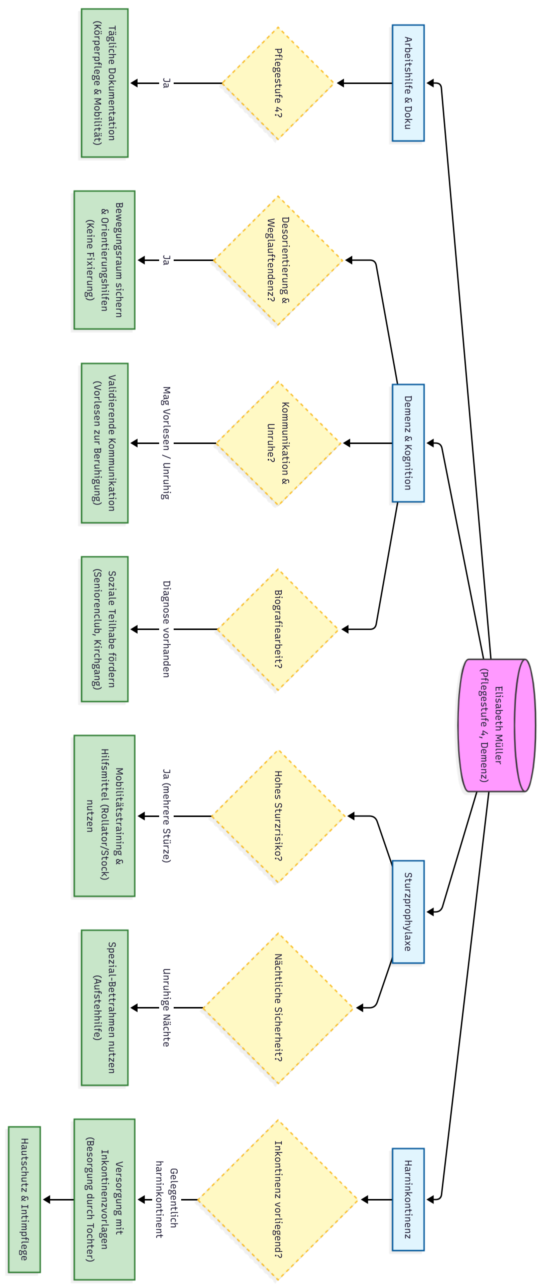
(a) Generated graph code using Gemini 3 Fast with English prompt and structured JSON input.

(b) Generated graph code using Gemini 3 Fast with english promt and and unstructured full text input.

Figure 10: Bigger display of Figure 4 and Figure 5



(a) Generated graph code using Gemini 3 Fast with German prompt and unstructured full text input.



(b) Generated graph code using Gemini 3 Pro with German prompt and unstructured full text input.

Figure 11: Display of the difference between Gemini 3 Pro and Gemini 3 Fast.

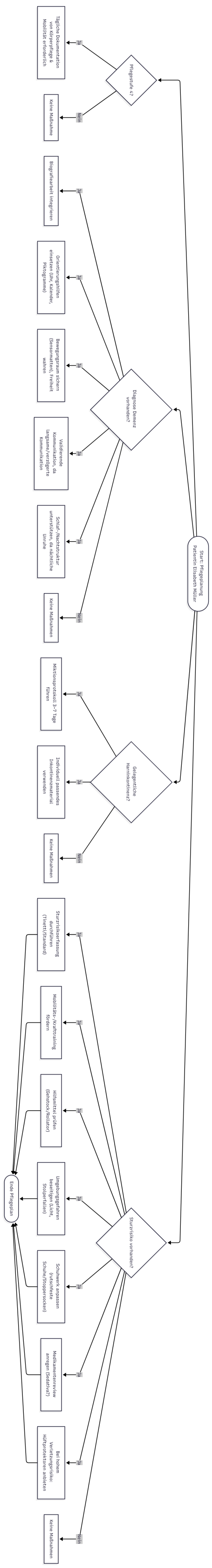


Figure 12: Generated graph code using GPT-5 with German prompt and structured logic input.