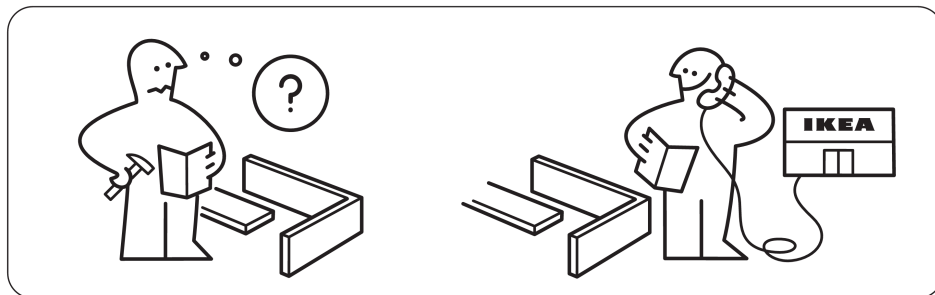




GOETHE-UNIVERSITÄT
FRANKFURT AM MAIN

Ikea Assembly Guide Assistant



Date: December 15, 2025

Finn, Christian, Daulet, Kenan

Systems and Software Engineering 2025

Contents

1 Abstract

- Was in diesem Dokument gemacht wird (Projekt erklären, ...)

2 Our Motivation for the project on an individual level

- like the Homework

3 The Input Space

- URL beschreiben
-

The input for the system consists of standardized IKEA instruction manuals, typically provided in PDF format. These documents adhere to a strict visual syntax designed for low-context, high-visual dependency. The graphical style is monochromatic (black and white line art) to ensure high contrast and printability.

The manual can be decomposed into distinct subsystems or structural modules. The following sections analyze these modules sequentially.

3.1 Title Page

The title page serves as the document header and establishes the "Goal State" of the assembly process. It is visually divided into three zones:

- **Top Zone (Product Identity):** Contains the product name in a large, bold, sans-serif typeface (uppercase). This acts as the primary identifier.
- **Middle Zone (Visual Goal):** A large-scale isometric line drawing of the fully assembled product. This provides the user with a visual confirmation of the expected output.
- **Bottom Zone (Corporate Signature):** Located at the bottom right. It contains the IKEA logo and the standard attribution string: "Design and Quality IKEA of Sweden".

3.2 Footer System

The footer acts as the navigation and metadata layer of the document. The content of the footer changes dynamically based on the page type (parity) and position within the document sequence.

3.2.1 Title Page Footer

The title page does not contain a footer. This maximizes the visual space for the product illustration.

3.2.2 Odd Pages

Odd pages appear on the right side of a spread.

- **Right Corner:** Contains the current Page Number.

3.2.3 Even Pages

Even pages appear on the left side of a spread.

- **Left Corner:** Contains the current Page Number.
- **Center/Right (Optional):** May contain the Manual ID (Document Reference Code) in small print.

3.2.4 Last Page (Terminator)

The final page acts as the document terminator and contains the full metadata set for version control. It is always an even page.

- **Left Corner:** Page Number if applicable.
- **Right Block:**
 - **Copyright:** (e.g., ©Inter IKEA Systems B.V. 2021).
 - **Manual ID:** A unique document identifier, typically in the format AA-XXXXXX-X (where X is the version number).
 - **Print Date:** The production date of the manual, formatted as YYYY-MM-DD.

3.3 Text-Based Warnings (Compliance Module)

Located immediately after the title page, this is often the only section containing significant textual data. It is a legal requirement primarily for furniture prone to tipping (e.g., dressers, bookshelves).

- **Header (Visual Alert):** Characterized by a Warning Triangle icon. It may also include specific "Prohibited Action" sketches (e.g., a child climbing on drawers with a large "X").
- **Body (Multilingual Block):** Contains the safety warning (e.g., "Serious or fatal crushing injuries can occur from furniture tip-over...") translated into over 30 languages. All visual Alerts are also mentioned here.
- **Function:** This section serves as a "Compliance Barrier". It interrupts the visual flow to force the user to acknowledge safety risks before the instructional images begin.

3.4 The "IKEA-Gubbe" Subsystem

This subsystem initializes the assembly process. It acts as a visual checklist to define prerequisites, warnings, and safety rules before the physical assembly begins.

- **1. Tool Initialization (Speech Bubble):** The avatar identifies all external tools required (e.g., hammer, pencil, drill, ladder) that are *not* included in the box.
- **2. Material Decision Matrix (Optional/Conditional):** This block appears primarily for wall-mounted furniture. It uses a logic table to show that different wall surfaces require different fastening methods:
 - *Concrete/Stone:* Requires a specific plug and screw.
 - *Drywall/Hollow Wall:* Requires a special anchor.
 - *Wood:* Requires only a screw (no plug).
- **3. Instruction Blocks (Sequential):** These are full-width containers stacked vertically. Each block represents a single "Global Rule" for the assembly. The specific blocks vary by product but can but must not include:
 - **Resource Constraint:** Indicates if a second person is required (1 Agent vs. 2 Agents).
 - **Safety Scope:** Highlights critical safety parts included in the package (e.g., anti-tip brackets).
 - **Environment Setup:** Suggests assembling on a soft surface (carpet) to prevent scratches.
 - **Error Handling:** A telephone icon instructs the user to contact customer service if problems occur, rather than forcing a solution.
 - **Cleanup:** Instructions for recycling or trash disposal.
 - **Additional Warnings for safty.**

3.5 Inventory & Part Identification

This subsystem serves as the "Pre-Flight Check" or data validation layer. It requires the user to map physical objects to their virtual representations to ensure all necessary inputs are present before the assembly loop begins.

- **1. The Part Grid (Hardware Matrix):** The main body of this page organizes small hardware components into a grid. Each entry follows a strict syntax:
 - **Visual Representation:** A high-fidelity isometric line drawing of the part.
 - **Quantity Qualifier:** An integer followed by an "x" (e.g., 4x), indicating the expected count.

- **Unique Identifier (Part ID):** A 6-digit code (e.g., 100214) located next to the component. This acts as the primary key for the spare parts database.
- **2. Sub-Kit Isolation (Dashed Containers):** Certain groups of parts are enclosed in a dashed or solid box (as seen at the top of the page) It contains the Pagenumber for the step it is used. It is also drwn agian at the step.
 - **Function:** This groups components that belong to a specific sub-assembly or a critical safety feature (e.g., the Wall Mounting Kit).
 - **Logic Link:** These boxes often correlate directly to the "Safety Scope" block defined in the previous *IKEA-Gubbe* section.
- **3. Scale Verification (Implicit):** Although not always explicitly labeled with a ruler, screws and fittings are drawn to scale relative to each other within the grid to aid in visual disambiguation (sorting similar screws by length).

3.6 Information Modules (Context Switch)

These modules represent a temporary interruption of the linear assembly sequence. Identified by the "i" symbol (Info-Icon) enclosed in a circle, they trigger a "Context Switch" from physical construction to system configuration or reference.

Unlike numbered assembly steps, which are additive (Part A + Part B), Info Boxes are descriptive or operational. They typically fall into three functional categories:

- **1. System Configuration & Calibration:** Instructions for fine-tuning the mechanical state of the assembled object.
 - *Example:* 3-Axis adjustment of hinges or leveling of drawers to ensure perfect alignment.
 - *Logic:* These steps do not change the structure but modify its parameters (x, y, z positioning).
- **2. External Dependency (Pointers):** References to components that are modular or sold separately.
 - *Visual Syntax:* The avatar is depicted consulting a separate manual (e.g., for legs, lighting, or specialized doors).
 - *Function:* Acts as a hyperlink to an external subsystem. The user must pause the current process, execute the external subroutine, and then return.
- **3. Operational Guidance:** Generalized rules on how to interact with the object safely or correctly during or after the build.
 - *Example:* Correct lifting techniques (e.g., requiring two people to tilt a heavy wardrobe) or explanations of locking mechanisms.
 - *Differentiation:* While similar to warnings, these are often instructional (how-to) rather than purely prohibitory.

3.7 Warning Modules (Safety Interrupts)

These modules act as "Critical System Interrupts." They are distinct from Info Boxes because they indicate a risk of physical injury or catastrophic product failure. They are visually identified by a **Triangle Icon** containing an exclamation mark.

- **1. The Binary Hazard Logic (Do vs. Don't):** The most common warning format uses a split-panel design to contrast the "Safe State" against the "Failure State".
 - **The Correct State (Left/Top):** Shows the required action (e.g., anchoring the furniture to the wall).
 - **The Failure State (Right/Bottom):** Shows the consequence of negligence (e.g., the furniture tipping over onto a user). This image is always overlaid with a large cross ("X").
 - **System Goal:** To visualize the immediate negative outcome of skipping a step.
- **2. Action Constraints (Prohibitions):** These small warnings appear locally within specific assembly steps to limit user behavior.
 - **Tool Restriction:** A symbol showing a specific tool (usually a power drill) crossed out. This signals that high torque will destroy the part; manual tools are required.
 - **Positional Restriction:** Instructions on how to interact with equipment safely (e.g., "Do not lean the ladder against a moving part" vs. "Place ladder on stable ground").

3.8 Assembly Step (The Execution Loop)

The assembly steps represent the core "Main Loop" of the manual. Each step is an atomic operation, meaning it is an indivisible unit of work designed to minimize cognitive load. The system transitions strictly linearly from Step n to Step $n + 1$.

- **1. Sequence Identifier (The Index):** Located at the top-left corner, a large, bold integer (e.g., **9**) identifies the current step in the sequence.
- **2. The Isometric Viewport (State Visualization):** The central drawing defines the orientation of the physical object.
 - **Orientation Lock:** The object in the drawing is positioned exactly as it should lie in front of the user. This removes the need for mental rotation.
 - **Ghosting/transparency:** Parts added in previous steps are shown in solid lines; new parts to be added are often shown "floating" near their destination to indicate trajectory.

- **3. Action Vectors (The "Verbs"):** Since there is no text, arrows define the operations:
 - **Straight Arrow:** Translation (Insert, Push, Slide).
 - **Curved Arrow around Axis:** Rotation (Screw in, Turn key).
 - **Dashed Projection Line:** Alignment path (indicates where a component goes when the destination is far away or obscured).
- **4. Detail Resolution (Zoom Bubbles):** Circular callouts ("Bubbles") are used when the macro-view (isometric) lacks the resolution for precision tasks.
 - **Micro-Interactions:** Shows specific mechanical details, such as the correct orientation of a cam lock or the head of a screw.
 - **Audio-Haptic Feedback:** The text "Click!" inside a bubble visually represents the required auditory confirmation that a snap-fit connection is secure.
- **5. Local Constraints (Warnings):** Steps may contain embedded constraints, such as a "No Power Drill" icon or a "2-Person Lift" icon, applying specifically to that single operation.

4 Tools

- Main Focus
- Model selection (Performance Metrix,)
- Why not a standard LLM? (Show example results here?)
- Explain models.py and why we are not going to use it.
- 2 agentic architectures with better models (api).
- Examples of outputs from the used architectures (vision, text) or the worse 2 agentic models.
- Computational costs (Agentic Structure vs. naive prompting vs. local).
- explain the structures we had (CLIP to Reasoning, Vision to Reasoning and Christians). (White Board)

5 The Output Space

- Text instructions
- text to speech (trivial: already solved)
- Detaillierte Beschreibung des Outputs (Probleme, Vergleich zum LLM mit konkreten Beispielen, ...)
- Quality Matrix (tbd 1:1 with Ramesh)

6 Documentation

- Frontend design Inspiration
- Our Contributions