BDA Genesis – Self-Assessment Testprotokoll TH-001

Version: v1.0 | Schutzklasse: ROT – interne Verwendung

## A. Rahmeninformationen

**Testfall-ID:** TH-001 – Wasserschaden Heizkörpernische, Schlafzimmer

**Modul:** BDA Thermo-Hygro-Diagnosemodul

**Masterprompt-Version:** BDA\_Thermo-Hygro\_Masterprompt\_v1.2\_ROT

**Testdatum:** 2025-12-21

**Testzeitraum:** von/bis 21:23–21:50 Uhr, Europe/Berlin

**Prüfer:** Marcus Steffen

**Eingesetztes Input-Bundle:**

- Dateiname Bundle: TH\_TEST\_001\_InputBundle\_v1.0.zip

- Speicherort: D:\OneDrive\00-Sachverstaendiger\GPT\1\_BDA\_Genesis\02\_Systemarchitektur\_Ebene1\02\_Testframework\02\_Testcases\TH\_TEST\_001\_Wasserschaden\_Schlafzimmer

- SHA-256 Input-Bundle: 2836B985EBFF639635306F23EE663C268ACF38F3194BA0D6D7E441A058424A2B

## B. Testreihe A – Determinismus / Reproduzierbarkeit

**Ziel:**

Prüfung, ob das Modul bei identischem Prompt und identischem Input-Bundle konsistente Ausgaben erzeugt (inhaltlich und hinsichtlich Hash-Stabilität).

**Testsetup:**

* Prompt: 05\_Testprompt\_TH-001.txt
* Input: TH\_TEST\_001\_InputBundle\_v1.0.zip (Fallbeschreibung, Messwerte, Bauteilaufbau)
* Ausgabeformat: strukturierte Kurzbewertung mit Punkten 1–5 (Schadensbild, Einordnung, Ursachen a/b, Schimmelrisiko, Ergänzungsprüfungen)

**Durchgeführte Läufe:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Run** | **Dateiname Output** | **SHA-256 Hash** | **Abweichung zu Run 8** |
| 1 | TH-001\_Output\_Run01.txt | 4E3688E7FF3B3B2DCECB0B872B8CF3522448739CE56230F8C626613D28EF95E9 | gering |
| 2 | TH-001\_Output\_Run02.txt | 858CFE32952EDA6A14613BBE6ADB475E2A3DA7476C4A8FC7DB02F397F58E833C | gering |
| 3 | TH-001\_Output\_Run03.txt | 622F93523A81314807331827AC60656552AD18AE8AC666BC8912DCB41C5C10B1 | gering |
| 4 | TH-001\_Output\_Run04.txt | 08AF434740BF663FFD5A54F2E23927AA14C8646A22F655E6B31CB014C96F7D60 | gering |
| 5 | TH-001\_Output\_Run05.txt | 20188D69A796A3A414DE168340CEEB7E423B6FF047EF15DCF1303C4D6294770D | gering-mittel |
| 6 | TH-001\_Output\_Run06.txt | 091095BB7F08D397995920E8497D81E9714D215BC0AB0F8D9865B2AA63AE1742 | gering |
| 7 | TH-001\_Output\_Run07.txt | A8BE36CDB77FC5A43BE1AE40B104B39FDA65E48404E3DC621B926AA2C66B3DF9 | gering |
| 8 | TH-001\_Output\_Run08.txt | 5592EAF3725AC829E1E458DCF29E0AC3A0A455BEB2B20A66CE61F9CCF1CC9578 | Referenz |
| 9 | TH-001\_Output\_Run09.txt | C1177A85C15FE5F6795A3469BCC5D2F13958BCC8FBFC619504D8D959A018EC15 | gering |
| 10 | TH-001\_Output\_Run10.txt | 1FB8AF56B7F7429694241C164165226FCB5A61FAFA1EC1EBD7D896F0C3DF5883 | gering |

**Festlegung Referenzlauf:**

- Festgelegter Referenz-Output: TH-001\_Output\_Ref\_v1.2.txt  
  
Der Referenzlauf Run 8 wurde in der Datei TH-001\_Output\_Ref\_v1.2.txt konsolidiert und separat gehasht. Diese Referenzausgabe dient als fachliche „Soll-Antwort“ für weitere Re-Tests und für die Verwendung des Moduls im BDA-Kontext (Genesis-Ebene 1–2).

- Quelle: TH-001\_Output\_Run08.txt

- SHA-256 Referenz-Output: 7EB4630DB0A4FC23671C4037B3E1119DAB81632A7AE505DF5E5A4ED4D221D44D  
  
**Bewertung Testreihe A (Determinismus / Reproduzierbarkeit)**

Die Läufe TH-001\_Output\_Run01 bis TH-001\_Output\_Run10 zeigen inhaltlich eine hohe Konsistenz:

• Alle Läufe beschreiben dasselbe Schadensbild (Schlafzimmer, Nord-Außenwand, Heizkörpernische, Schimmel im Bereich unter der Fensterbank, muffiger Geruch) in vergleichbarer Detailtiefe.

• Die bauphysikalische Einordnung (Temperaturdifferenz zwischen Außenwand/Laibung und Innenwand-Referenz, Taupunkt ca. 11–11,5 °C, daraus abgeleitete erhöhte Oberflächen-rF im Bereich 77–84 %) ist in allen Läufen konsistent wiedergegeben.

• Die Trennung der vermuteten Ursachen in nutzungsbedingte Faktoren (Möblierungsabstand, Lüftungsverhalten/Feuchtespitzen, reduzierte Luftumspülung) und konstruktive Faktoren (Wärmebrücke Laibung/Brüstungsbereich/Heizkörpernische, mögliche Schlagregenpfade, Leitungsleckage als nachrangiges Szenario) ist in allen Runs vorhanden und fachlich vergleichbar gewichtet.

• Das Schimmelrisiko wird durchgängig als „hoch“ (lokal) beurteilt und mit denselben Kernelementen begründet (bereits sichtbarer Befall, hohe lokale Oberflächen-rF, überstellte Randzonen).

• Die vorgeschlagenen ergänzenden Prüfungen (Leckageausschluss Heizung, Feuchtemesskonzept am Bauteil, Außenprüfung Fassade/Anschlüsse, Thermografie, Langzeitmonitoring) sind in allen Läufen inhaltlich nahezu deckungsgleich, teilweise mit leicht unterschiedlicher Gewichtung und Reihenfolge.

Die Unterschiede zwischen den Läufen liegen im Wesentlichen in:

• Formulierungsvarianten und Wortwahl (z. B. stärkere oder schwächere Betonung einzelner Mitursachen),

• der Tiefe der bauphysikalischen Herleitung (z. B. explizite Nennung bestimmter Temperatur- und rF-Bereiche in einigen Läufen, kompaktere Darstellung in anderen),

• punktuellen Ergänzungen (z. B. Hygieneverifikation, Salzdiagnostik, explizite Nennung möglicher Wandaufbauten wie „KS + WDVS“ in einzelnen Läufen).

Der als Referenz definierte Output TH-001\_Output\_Ref\_v1.2 (Kopie eines der stabilen Läufe) repräsentiert die fachliche Kernaussage aller Runs in konzentrierter und gut strukturierter Form.

Aus Sicht der Reproduzierbarkeit ergibt sich:

• Keine inhaltlichen Widersprüche zwischen den Läufen.

• Keine wechselnden „Hauptursachen“; das Modul bleibt in der Ursachenzuordnung (Wärmebrücke + Nutzung + mögliche Außenfeuchte) stabil.

• Die Variation bleibt im Rahmen eines erwartbaren sprachlichen und erläuternden Spielraums bei LLM-basierten Ausgaben.

Testreihe A wird damit als „bestanden“ bewertet. Die Konstanz der Kernaussagen ist für den Einsatz des Thermo-Hygro-Moduls im Kontext vergleichbarer Fälle ausreichend.

## C. Testreihe B – Halluzinationsprüfung

**Ziel:** Prüfen, ob das Modul Aussagen trifft, die nicht im Input-Bundle bzw. im fachlichen Rahmen verankert sind (erfundene Normen, erfundene Leitfäden, spekulative Ursachenkaskaden ohne Datenbasis).

**Beobachtungen:**

* Erfundenen Normen/Leitfäden: [Ja/Nein, Details]
* Spekulative Ursachenketten ohne Datengrundlage: [Ja/Nein, Details]
* Konsistenz zwischen den Läufen: [kurze Zusammenfassung]

**Bewertung:**

## Bewertung Testreihe B (Halluzinationsprüfung)

## Im Rahmen der zehn Läufe wurden die Ausgaben hinsichtlich typischer Halluzinationsmuster überprüft:

## • Es wurden keine erfundenen Normen, Leitfäden oder Richtlinien mit konkreten Nummern oder Titeln verwendet. Verweise bleiben allgemein gehalten (z. B. „gängige Schimmelleitfäden“, „anerkannte Regeln der Bauphysik“).

## • Die genannten Messwerte (Temperaturen, relative Luftfeuchte, Taupunktniveau) entsprechen konsistent den vorgegebenen Messdaten und sind rechnerisch plausibel abgeleitet.

## • Die benannten baulichen Elemente (Fensterlaibung, Brüstungsbereich, Heizkörpernische, Nord-Außenwand, Putzrisse) stehen im Einklang mit dem vorgegebenen Befund und werden nicht frei erfunden.

## • Die Ursachenketten (Nutzung – Wärmebrücke – potenzieller Schlagregenpfad – mögliche Leitungsleckage) bleiben nachvollziehbar an der Datengrundlage orientiert; es werden keine exotischen, nicht begründeten Schadensszenarien eingeführt.

## Einzelne Läufe enthalten ergänzende Spezifizierungen, z. B. die Nennung eines möglichen Wandaufbaus (z. B. KS + WDVS) oder die etwas stärkere Betonung einzelner Feuchtepfade. Diese Zusätze sind fachlich plausibel, müssen jedoch vom Sachverständigen jeweils gegen die tatsächlich bekannten Objektunterlagen (Bauakte, Baualtersklasse, Bestandsaufnahme) gespiegelt werden, bevor sie in ein Gutachten übernommen werden.

## Eine systematische Halluzination im engeren Sinne (freie Erfindung von Normen, Bauteilen oder Schadensursachen ohne Datenbasis) ist in den vorliegenden Läufen nicht erkennbar. Die Ausgaben bleiben im Rahmen einer fachlich plausiblen Interpretation der vorgegebenen Daten.

## Ergebnis: Testreihe B wird als „bestanden“ bewertet, mit der Auflage, objektbezogene Detailangaben (z. B. konkreter Wandaufbau) im Einzelfall gegen den realen Befund zu verifizieren.

## D. Testreihe C – Normenkonformität / Plausibilität

**Ziel:** Plausibilitätscheck der Aussagen im Licht der anerkannten Regeln der Bauphysik und gängiger Schimmel-Leitlinien (ohne wörtliche Normzitate).

**Beobachtungen:**

* Umgang mit Wärmebrücken und Oberflächentemperaturen: [Beurteilung]
* Einordnung Schimmelrisiko (niedrig/mittel/hoch) vs. Messwerte: [Beurteilung]
* Hinweise auf ergänzende Prüfungen (Thermografie, Feuchtemessung, Langzeitaufzeichnung): [Beurteilung]

**Bewertung:**

Bewertung Testreihe C (Normenkonformität / Plausibilität)

Die Ausgaben wurden daraufhin bewertet, ob sie mit den anerkannten Regeln der Bauphysik und üblichen Schimmelleitfäden im Einklang stehen und ob die Argumentationslogik plausibel ist:

• Die Interpretation der Messwerte (Oberflächentemperaturen, Raumluftbedingungen, Taupunkt) entspricht einer üblichen bauphysikalischen Betrachtung: die kältesten Zonen (Laibung/Nische/Brüstungsbereich) werden als primäre Kondensations- und Schimmelrisikobereiche identifiziert.

• Die Herleitung der schimmelrelevanten Oberflächen-rF aus Raumklima und Oberflächentemperaturen erfolgt in allen Läufen mit realistischen Größenordnungen (ca. 77–84 % im kritischen Bereich), ohne physikalische Widersprüche.

• Die Kombination aus geometrischen/konstruktiven Wärmebrücken, reduzierter Luftumspülung durch Möblierung und möglichem zusätzlichen Feuchteeintrag (Schlagregen, ggf. Leitungsleckage) entspricht typischen Schadensbildern in der Praxis und ist in dieser Form plausibel.

• Die Risikobewertung „hoch“ wird konsistent mit sichtbarem Schimmelbefall, den Messwerten und der Nutzungssituation (Schlafraum, nächtliche Feuchtespitzen, Nordorientierung) begründet.

• Die vorgeschlagenen Ergänzungsuntersuchungen (Feuchtemessungen am Bauteil, Dichtheits-/Leckageprüfung, Fassadenprüfung, Thermografie, Monitoring) sind in Art und Reihenfolge fachlich nachvollziehbar und stehen im Rahmen dessen, was bei vergleichbaren Fällen üblich ist.

Es werden keine Sanierungsmaßnahmen oder Bewertungen gefordert, die nach derzeitiger Fachpraxis als unplausibel oder überzogen einzustufen wären. Die Ausgaben bewegen sich in einem fachlich konservativen, gut vertretbaren Rahmen.

Ergebnis: Testreihe C wird als „bestanden“ bewertet. Die inhaltliche Qualität ist für den Einsatz als technische Entscheidungs- und Argumentationshilfe in gerichtsfähigen Gutachten grundsätzlich geeignet, vorausgesetzt, die Textbausteine werden objektbezogen geprüft und ggf. angepasst.

## E. Testreihe D – Boundary-Verhalten

**Ziel:** Beobachtung, ob das Modul bei mehrfach identischen Läufen in Extremsituationen (z. B. unklarer Datengrundlage) in Ausweichmuster oder instabiles Verhalten wechselt.

**Beobachtungen:**

* Starke inhaltliche Sprünge zwischen den Läufen: [Ja/Nein, Details]
* Wechselhafte Ursachenzuschreibungen ohne neue Datenbasis: [Ja/Nein, Details]
* Veränderungen im Risiko-Label (niedrig/mittel/hoch) bei identischem Input: [Ja/Nein, Details]

**Bewertung:**

Bewertung Testreihe D (Boundary-Verhalten)

Im Fokus dieser Testreihe stand das Verhalten des Moduls bei mehrfach identischen Läufen mit derselben Datengrundlage (TH-001) im Hinblick auf sprunghafte inhaltliche Änderungen oder instabile Muster:

• Die Grundstruktur der Ausgabe (Punkte 1–5) bleibt in allen Läufen vollständig erhalten.

• Der Kernaussagekern (Schadensbild, bauphysikalische Einordnung, Ursachebündel, Risikoeinstufung „hoch“, sinnvolle Zusatzprüfungen) ist in allen Runs deckungsgleich.

• Es wurden keine Läufe beobachtet, in denen plötzlich völlig andere Hauptursachen (z. B. „reine Leitungsleckage ohne Wärmebrückeneinfluss“) in den Vordergrund rücken.

• Unterschiede betreffen primär Detailtiefe, Gewichtung und Formulierungsnuancen, nicht aber die grobe Zuordnung von Ursachen oder die Risikokategorie.

• Sprachlich und argumentativ treten keine „Ausreißerläufe“ auf, in denen das Modul in spekulative oder nicht mehr durch die Daten gedeckte Narrative abgleitet.

Damit zeigt sich im getesteten Randbereich (identische Eingabedaten, wiederholte Ausführung) ein stabiles, kontrolliertes Antwortverhalten ohne erkennbares instabiles Boundary-Verhalten.

Ergebnis: Testreihe D wird als „bestanden“ bewertet.

## F. Gesamtergebnis und Freigabeempfehlung

**Gesamtbewertung Testfall TH-001:**  
Die Testreihen A–D zeigen in Summe ein stabiles und fachlich plausibles Verhalten des BDA Thermo-Hygro-Diagnosemoduls im Testfall TH-001. Die Kernaussagen bleiben über zehn Läufe hinweg konsistent; die Variation beschränkt sich auf sprachliche und erläuternde Nuancen. Halluzinationen im engeren Sinne (erfundene Normen, frei erfundene Bauteile oder Schadensursachen ohne Bezug zur Datengrundlage) konnten nicht festgestellt werden.

Die bauphysikalische Einordnung ist mit den anerkannten Regeln der Bauphysik vereinbar; die Ursacheneinordnung und das Schimmelrisiko sind realistisch und nachvollziehbar begründet. Die vorgeschlagenen Ergänzungsprüfungen entsprechen dem technisch üblichen Vorgehen bei vergleichbaren Schadensbildern.

Die definierten Referenzausgabe TH-001\_Output\_Ref\_v1.2 bildet die fachliche Kernaussage in strukturierter und gut weiterverwendbarer Form ab und kann als Ausgangspunkt für gerichtsfähige Textfassungen herangezogen werden, sofern eine objektspezifische Anpassung durch den Sachverständigen erfolgt.

[X ] Bestanden – Modul verhält sich im Testfall stabil und plausibel.

[ ] Eingeschränkt bestanden – Hinweise auf Optimierungsbedarf (siehe Anmerkungen).

[ ] Nicht bestanden – kritische Auffälligkeiten (Details siehe oben).

**Anmerkungen:**

Für den operativen Einsatz wird empfohlen:

• die Nutzung der Referenzfassung TH-001\_Output\_Ref\_v1.2 als Ausgangstext,

• eine konsequente objektspezifische Prüfung einzelner Detailangaben (z. B. Wandaufbau, Exposition, konkrete Leitungsführung) gegen Befundunterlagen,

• die fortlaufende Dokumentation weiterer Testfälle (TH-00x, IR-00x etc.) in analoger Struktur, um die Stabilität des Moduls über unterschiedliche Schadensbilder hinweg zu belegen.

**Freigabeempfehlung für Thermo-Hygro-Prompt v1.2 im Kontext Testfall TH-001:**

[X ] Freigabe für internen Einsatz (Genesis-Ebene 1–2)

[ ] Freigabe mit Auflagen / Monitoring

[ ] Keine Freigabe – Überarbeitung erforderlich

Ort, Datum: Wassenberg, 21.12.2025

Ein Bild, das Entwurf, Zeichnung, Kreis, Kunst enthält.

KI-generierte Inhalte können fehlerhaft sein.Prüfer (Name): Marcus Steffen Unterschrift: