Japheth Dangiwa Jerry

Below is an **opponent-style review** of **Dangiwa.pdf** using the **KJU opponent's evaluation criteria and scoring scale** in **opponensi_lap_en.pdf**.

---

0) Quick thesis snapshot (what the work claims to do)

The thesis examines the **"limits and opportunities of LLM agents"** via a **case study**: building and experimentally evaluating a **Dance Class Management and Analytics System (DCMAS)** with **Experiments #1–#7** covering trend analysis, missing data, prompt sensitivity, temporal instability, and goal/value ambiguity.

It explicitly states the system uses **synthetic datasets** for privacy-preserving testing.

---

1) Use of AI & suspected plagiarism (risk-focused assessment)

1.1 Declared / visible AI use (high)

- The thesis includes a **"Derived LLM Prompt (used for experimentation)"** and an annex titled "**Relevant LLM-Conversations with Full Texts**", indicating **direct, intentional use of an LLM as part of the research method**.
- Multiple chapters describe **LLM prompt design, prompt sensitivity, and non-deterministic outputs** as core experimental content.

✅ **Positive:** AI use is not hidden; it is framed as an object of study and experimentation.

⚠️ **Concern:** The thesis should clearly distinguish **(a) AI used as the research subject/tool** vs **(b) AI used to draft the thesis text** (authorship transparency). The provided text does not clearly document AI assistance in writing, only in experimentation.

1.2 Plagiarism / originality risk indicators (medium, not proven)

I cannot "prove" plagiarism from PDFs alone; I can only flag **risk indicators** that warrant a formal similarity check.

Indicators observed:

- Several definitions are pulled from **Wikipedia** (e.g., management information system, structured data, descriptive analytics, human-in-the-loop), which is not plagiarism *if correctly cited*, but it is academically weak and increases similarity risk because these texts are widely reused.
- Large portions of the prose have a **highly uniform, template-like style** ("This experiment demonstrates…", repeated structure), which is **consistent with AI-assisted drafting** *or* disciplined technical writing; this is a **style-based indicator only**, not evidence.
- There is a peculiar **Section 2.6** containing what looks like a **curriculum/course mapping** ("Introduction to Algorithms… Operating Systems…") embedded inside the Literature chapter outline, which is atypical and may indicate patchwork editing or "compliance insertion."

Recommendation (standard academic practice):

- Run an institutional similarity report (Turnitin/Urkund/etc.) and require an **AI-use declaration** describing:
  a. which tools were used,
  b. for what tasks (experiments vs writing),
  c. what was human-authored,
  d. how outputs were verified.

This is particularly relevant because the thesis studies LLMs and includes LLM transcripts.

2) Scored evaluation (max 200 points total)

> Scoring categories follow the opponent form: **Topic/Objectives (40)**, **Literature (40)**, **Own work (60)**, **Structure/style/formal (40)**, **Overall (20)**.
>
> Grade bands: **Fail 0–119**, **Satisfactory 120–139**, **Average 140–159**, **Good 160–179**, **Excellent 180–200**.

2.1 Topic and Objectives (max 40) — 30/40

Strengths

- Topic is relevant and current: **LLMs in analytics/management decision support** and their limitations in practice.
- Research question and objectives are clearly stated (limits of replacement; experiments #1–#7).

Weaknesses

- The framing sometimes shifts toward broad general claims ("replace sizable portion…") with only rough estimation (hours, € savings) and without robust justification.

Rationale: Clear objectives + relevant problem; minor overclaiming risk.

2.2 Review of the Literature (max 40) — 20/40

Strengths

- Includes recognizable foundational sources (Provost & Fawcett; Davenport & Harris; Brown et al.; ISO/IEC).
- Literature is aligned with thesis themes: data-driven decision-making, DSS, LLM limitations.

Weaknesses

- The "Literature" chapter is **thin and mostly high-level**, with limited synthesis (few competing viewpoints, limited depth on DSS/analytics evaluation, limited discussion of methodological standards for case studies).
- Heavy reliance on **Wikipedia definitions** for key terms is academically weak for a thesis-level literature review.
- The presence of **Section 2.6 course-like subsections** inside Literature undermines scholarly focus and coherence.

Rationale: Some credible sources exist, but analysis/synthesis and academic depth are insufficient.

2.3 Presentation of the Author's Own Work (max 60) — 34/60

Strengths

- The experimental structure (#1–#7) is systematic and escalates complexity (missing data, prompt sensitivity, temporal instability, goal ambiguity).
- The thesis emphasizes **human oversight, validation, responsibility boundaries**, which is conceptually strong and consistent with the experiments.
- It claims transparency via full prompt/output transcripts in annexes (good practice if complete).

Weaknesses

- The work reads more like a **conceptual demonstration** than a rigorously validated system study: limited visible technical detail about implementation, datasets, KPI definitions, and evaluation metrics (in the provided content).
- Economic/time-saving estimates are **rough** (e.g., hours/week and €/year) and not backed by a clear measurement protocol.
- Without a reproducible method section (exact model settings, temperature, sampling, repetition count, scoring rubric),

claims about instability may be correct but are not “research-grade” demonstrated.

Rationale: Good conceptual design and experiment logic; evidence and methodological rigor appear incomplete.

2.4 Structure, Style, and Formal Aspects (max 40) — 22/40

Strengths

- Clear top-level structure (Intro → Literature → Case Study → Methodology → AI Integration → Oversight → Discussion → Conclusion).

Weaknesses

- Noticeable formal issues: duplicated supervisor name, odd spacing/encoding in institutional names, and general presentation noise reduce professionalism.
- The inclusion of the course-like content in Chapter 2 harms logical proportion and academic style.
- Mixed citation practice (book-style references vs raw Wikipedia URLs) weakens formal consistency.

Rationale: Understandable structure, but formal/academic presentation needs tightening.

2.5 Overall Impression (max 20) — 13/20

The thesis is **topically relevant** and has a **coherent experiment narrative**, but it needs stronger **methodological rigor**, a more scholarly **literature synthesis**, and improved **formal presentation** to meet a stronger academic standard.

---

3) Total score and grade band

	Criterion	Max	Score
1	Topic & Objectives	40	**30**
2	Literature Review	40	**20**
3	Author’s Own Work	60	**34**
4	Structure/Style/Formal	40	**22**
5	Overall Impression	20	**13**
6	**TOTAL**	**200**	**119 / 200**

Result by KJU scale: 119 points = Fail (0–119).

Conditional note (important)

If the student can (1) document originality and (2) substantially strengthen the methodological detail (model settings, replication protocol, scoring rubric, dataset specification) and (3) clean formal issues, the work could plausibly move into **Satisfactory (120–139)** territory.

---

4) High-impact improvement list (what to fix first)

1. **Methodology hardening (must-have):** specify the LLM version, parameters (temperature), number of repeated runs, evaluation rubric, and how human judgments were recorded.
2. **Replace Wikipedia definitions** with peer-reviewed/standard references and synthesize literature into an argument (compare positions, identify gaps).
3. **Remove/relocate Section 2.6 course mapping** unless it is an explicit institutional requirement (then move to annex and explain purpose).
4. **Evidence of "system" work:** include architecture diagram, data schema, example KPI calculations, sample dashboards, or code repository description (even if private, describe components).
5. **AI transparency statement:** clearly declare what AI did in the research and whether it assisted in writing/editing.

---

5) Suggested defense (oral) questions (good for opponensi)

1. How many repeated runs did you do for Experiments #5–#6, with what parameters, and how did you quantify "instability"?
2. What is your exact KPI set in DCMAS, and how are KPIs computed from the dataset?
3. How did you validate that the synthetic data preserves real operational distributions (seasonality, churn, payment behavior)?
4. Where do you draw the boundary between "decision support" and "decision making" in system design, and how is this enforced technically?
5. What parts of the thesis text (if any) were AI-assisted, and how did you ensure originality and correctness?

If you want, I can also produce a **filled-in opponent form style narrative** (short "General summary" + paragraph per criterion + final recommendation) matching your university's typical wording—just tell me whether you prefer a **strict** or **constructive/upgrade-oriented** tone.