

CONNECTIONS UK 2026

AI Across the Wargame Cycle

What Two Years of Focused Exploration Actually Taught Me

COL (R) Robert Burks, Ph.D.

Naval Postgraduate School, Wargaming Center





Where We're Going

- Setting the Table
- The Lifecycle
- The Evidence
- The Warning
- The Proposition
- Wrap Up



Today's Focus is on how Wargaming Practitioners can come to grips with AI – but all wargamers can learn a thing or two today as well.

Brief Bio

Colonel (R) Robert Burks, Ph.D.

- 32 years of U.S. Army service
 - Over a decade as Enlisted Infantryman and Infantry & Quartermaster Officer
 - 17 years OR/Systems Analyst
- **50+ years of Professional & Hobby Gaming.**
- Currently
 - Associate Professor, Defense Analysis
 - Senior Wargame Designer NPS Wargaming Center
- Supporting DoD, Partners and Allies
 - **44 organizations, with 60+ workshops, to help grow their wargaming capability.**



Naval Postgraduate School
Wargaming Center
Senior Wargame Designer





What Is a Wargame?

An Act of Communication that simultaneously **explores complexity**, enhances understanding and **communication**, **stimulates creativity**, and **contributes to a unity of action**.

A dynamic representation of **conflict or competition** in which **people make decisions** and respond to the **consequences** of those decisions.



What About these ...

Crisis Simulation

Conflict Simulation

Crisis Management

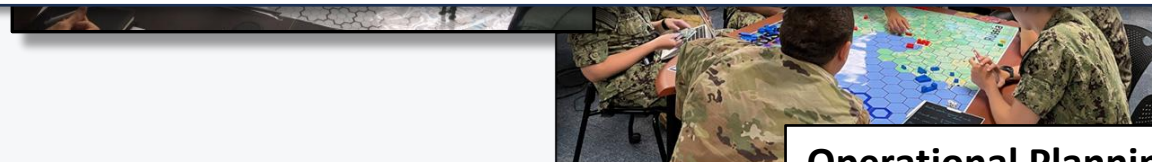
Policy Gaming

Serious Gaming

Peace and Conflict Exercises

Professional Gaming

... are they wargames?



Operational Planning



Why We Wargame

The goal of all games is to immerse the players in an environment with the required level of realism to permit the investigation of the human decision-making process.

Three broad purposes (**The Why**) of gaming:

- **Educational (Player Focused – conveying/impart knowledge)**
 - **Reinforce learning objectives** while exposing players (students) to situations that they are likely to encounter during their professional careers.
- **Experiential (Player Focused – conveying knowledge)**
 - Provide players practical experience during training that better prepares them for specific jobs or tasks.
- **Analytical (Sponsor Focused – extract knowledge)**
 - **Extract insights and findings** from players to inform organizational challenges and complex problems.



The Wrong Question — and the Right One



“Can AI design a wargame?”

The wrong threshold. It invites a yes/no verdict on a tool, not a judgment about a process — and it sets AI up as an easy replacement for the designer, facilitator, or analyst rather than a partner to them.



“Where can AI increase the capacity of a professional wargaming team, inside the lifecycle it already uses?”

Not a new AI wargaming methodology. AI augmenting the methodology practitioners already trust — tested against eight real games, not a slide deck of possibilities.

The failure mode that matters most across all eight games wasn't obvious error — it was plausible-sounding output that passed a fluency check but failed a logic check against history, feasibility, or doctrine.

We'll come back to that idea at the end — it's the thread that ties the evidence together.



THE FRAMEWORK

Anchor: The Analytic Wargaming Lifecycle

FIVE PHASES

↻ ITERATE



Where AI fits is different in every phase — and so is how much we should trust it.

NOT SPECULATION



Two Years. Eight Executed Games.

2

YEARS

8

EXECUTED GAMES

6→300+

PLAYERS PER GAME

+LLM-only

GAMES: AI AS ALL PLAYERS

THE GAMES

NWPAC × 2

300+ player US–Japan bilateral

Zefra

14-player seminar game

Stalingrad

6-player matrix game

Crimea Annexation

Educational campaign game

Grozny Grinder

Educational campaign game

Salang's Gauntlet

Educational campaign game

ÜBUNG

Educational campaign game

Plus: a structured synthetic-playtesting method (Strange Horizons) and a classified LLM adjudication effort in development.

Five slides from here: Stalingrad (the thread), Strange Horizons, Grozny Grinder, Zefra, and NWPAC — five distinct kinds of evidence, not one story repeated five times.

AI as a Questioning Partner, Not a Content Generator

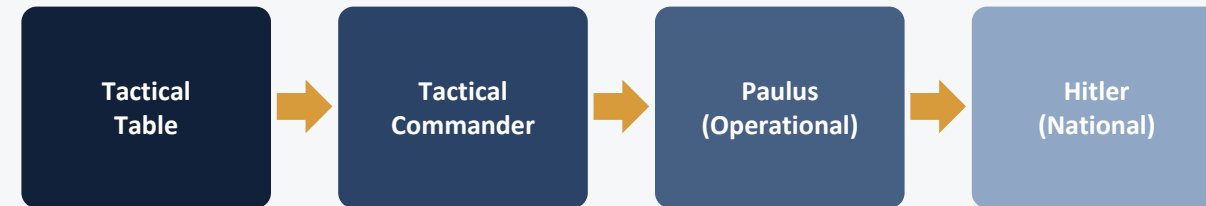
Stalingrad Decision Lab is our through-line: a matrix game spanning strategic, operational, and tactical command, developed with AI from initial framing through the AI-headquarters experiment.

What decision does this rule actually force?

What information reaches higher headquarters — and what happens if that report is wrong?

Does this rule contradict something we decided three iterations ago?

THE DESIGN CONSEQUENCE: INFORMATION AS A COMMAND MECHANIC



Each link filters and delays what the next level knows — the reporting chain became a game mechanic, not a UI convenience.

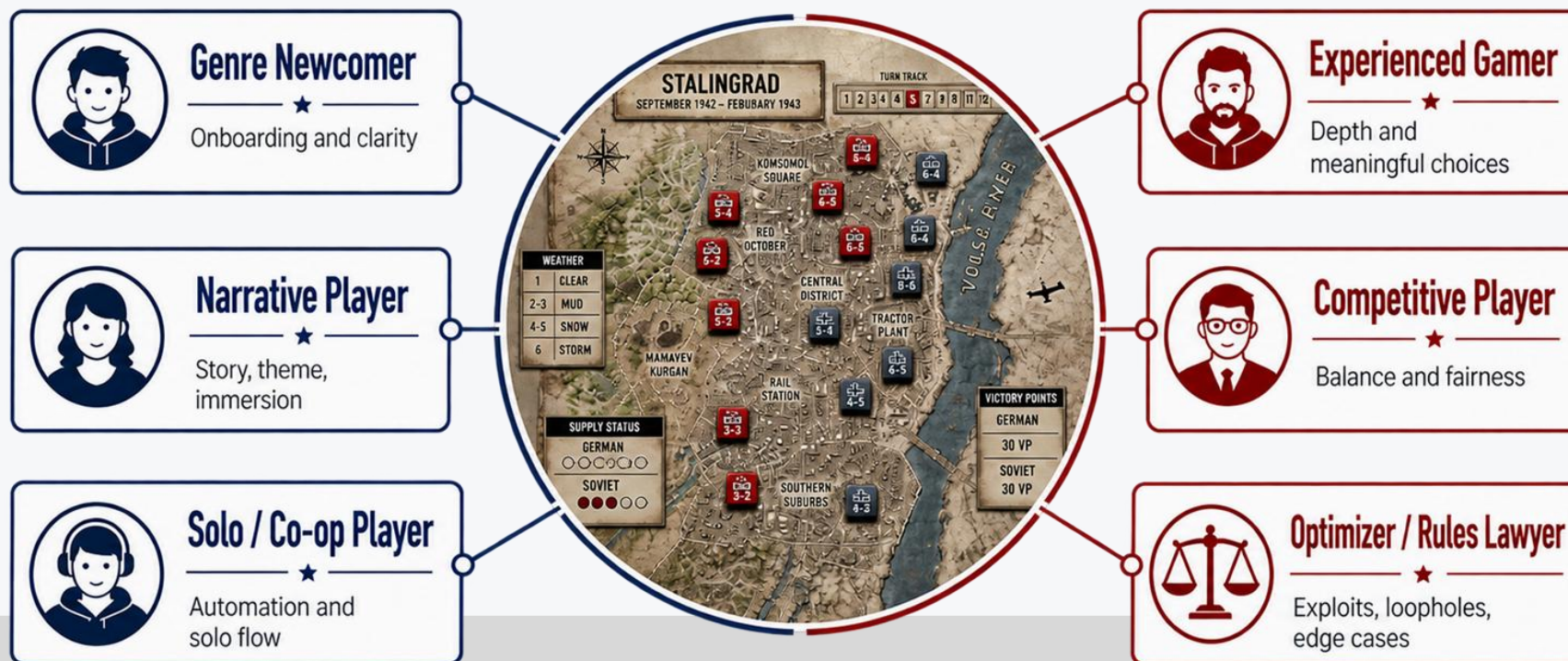
“Change who knows the tactical situation — not the tactical combat system itself.”

It all started with a prompt ... “For the workshop portion, I am thinking of using a game focused on Stalingrad that includes tactical, operational and strategic level players. Provide a frank assessment of the strength of the concept then provide a recommended updates/changes to make it a strong candidate.”

Give the Playtest a Mission, Not Just a Player

STRANGE HORIZONS — 8 ITERATIONS • 6 PURPOSE-BUILT PLAYTESTER PROFILES

Strange Horizons was tested with six purpose-built AI profiles, run through a Design-of-Experiments matrix (profile × crew × resource level × pacing) across iterative cycles.



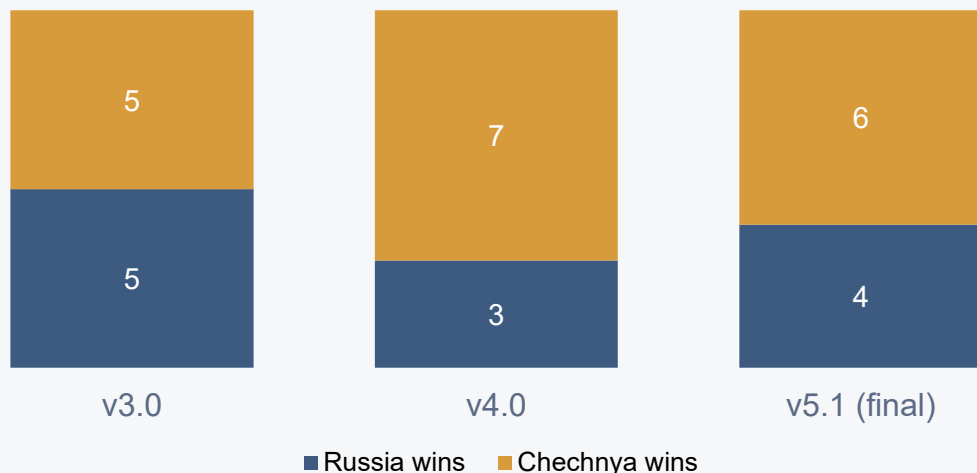
THE FINDING

The Optimizer found a dominant opening — rush Scene Control and the Signal Crate. The question wasn't "would a normal player do this?" It was "can the rules permit an intelligent player to?"

Watching a Design Converge, Version by Version

Grozny Grinder teaches why a stronger conventional force can still lose a war of political endurance to a weaker force that trades territory for time.

Result Split Across 10 Playtests per Version



Grozny Grinder: AI played all roles across three structured playtester rounds, each cycle re-tested through ten runs before the next rewrite (Versions – 3.0 – 5.1).

V1/2.0

Human Written – Believed to be achieving the learning objectives.

v3.0

5–5 split — but untrustworthy: rules friction, not design, was shaping outcomes.

v4.0

7–3 — clearer rules, but the historian judged it over-corrected against Russia.

v5.1

6–4 — and for the right reasons: Russia lost to corridors and time, not broken rules.

Zefra: Running the Same Seminar Game Three Ways

A 14-player UN/CJTF stabilization seminar game (Crisis in Zefra), run three times to directly compare human and AI performance in the same roles.

BASELINE

Round 1 — Human Players

Full human seminar game. Roles included national leaders, CJTF-Z command, NGOs, factions, and media.

AI AS ANALYST

Round 2 — AI-Assisted Analysis

Human players again — but AI captured player responses and produced the post-game analysis.

AI AS PLAYERS

Round 3 — AI as All Players

AI assumed every role in the game.
Outcome: AI performed comparably to our non-expert human players.

THE FINDING

The question wasn't "can AI play this game?" It was "can AI capture and analyze what humans decided, fast enough to matter?" Here, yes to both — but **only to the standard of a non-expert player**, not a subject-matter expert.

NWPAC: When Hundreds of Players Make Decisions at Once

Northwest Pacific (NWPAC) is NPS's CNO-directed, Title 10 U.S.–Japan bilateral wargame — executed in 2024 and 2025, each with 300+ participants across U.S. and Japanese forces.

Speed-of-relevancy adjudication support for kinetic operations — decisions at a pace a human control cell alone cannot sustain.

Post-analysis support across hundreds of simultaneous player decisions.

A classified LLM variant is in development for this effort — not detailed here.

300+

PARTICIPANTS PER ITERATION

2 EXECUTED ITERATIONS (2024, 2025)

U.S. Navy × Japan Maritime Self-Defense Force · CNO-directed Title 10 wargame series

Source: Naval Postgraduate School, NWPAC public program materials.

THE FINDING

This remains experimental, not operational. The bottleneck isn't whether AI can support the game — it's capturing hundreds of simultaneous player decisions into a structured format AI can process in the first place. At this scale, the sheer volume of parallel conversations exceeds what we can currently capture, let alone analyze in full.



THE WARNING

The Risk Isn't Obvious Error

It's output that passes a fluency check and fails a logic check — against historical accuracy, operational feasibility, or doctrinal soundness.

PASSES THE FLUENCY CHECK

- Confident, well-structured adjudication language
- Internally coherent scenario detail
- Consistent tone and military register

FAILS THE LOGIC CHECK

- Grozny v4.0: clean, teachable rules — still over-punished Russia until a historian caught it.
- Unsupported assumptions treated as fact
- Overconfident recommendations with no basis to check them against

Some risks a logic check won't catch at all: the physical weight of a rule at the table, and how two real players read each other. No amount of AI iteration substitutes for that.

Version drift, hallucinated detail, and "too-competent" synthetic staffs all share this shape — fluent, and wrong.



THE GOVERNING FRAMEWORK

“Can AI Do It?” Is the Wrong Test

The better question: what role is AI performing, and what authority have we given it?



Stalingrad has now touched nearly every point on this continuum — from rules discussion to the AI-headquarters experiment now under test.

LOW RISK

Rewriting a player aid or facilitator guide

MODERATE RISK

Proposing alternative mechanics — the designer still chooses

HIGH RISK

Synthetic playtesting used as evidence about real decision-makers



AI-Enabled Wargaming, Not AI-Generated Wargaming



We still need humans to determine the problem, judge context, adjudicate ambiguity, and decide what evidence means. AI gives a small wargaming team something it has rarely had: inexpensive additional capacity across the entire lifecycle.

Thank You

Questions welcome.

THIS AFTERNOON

Stalingrad Decision Lab — Play, Prompt, Validate

A hands-on demonstration of LLM-supported analytical wargaming, set during the November 1942 Stalingrad encirclement crisis. Play, prompt, and apply a validation checklist at four controlled points in the game.

