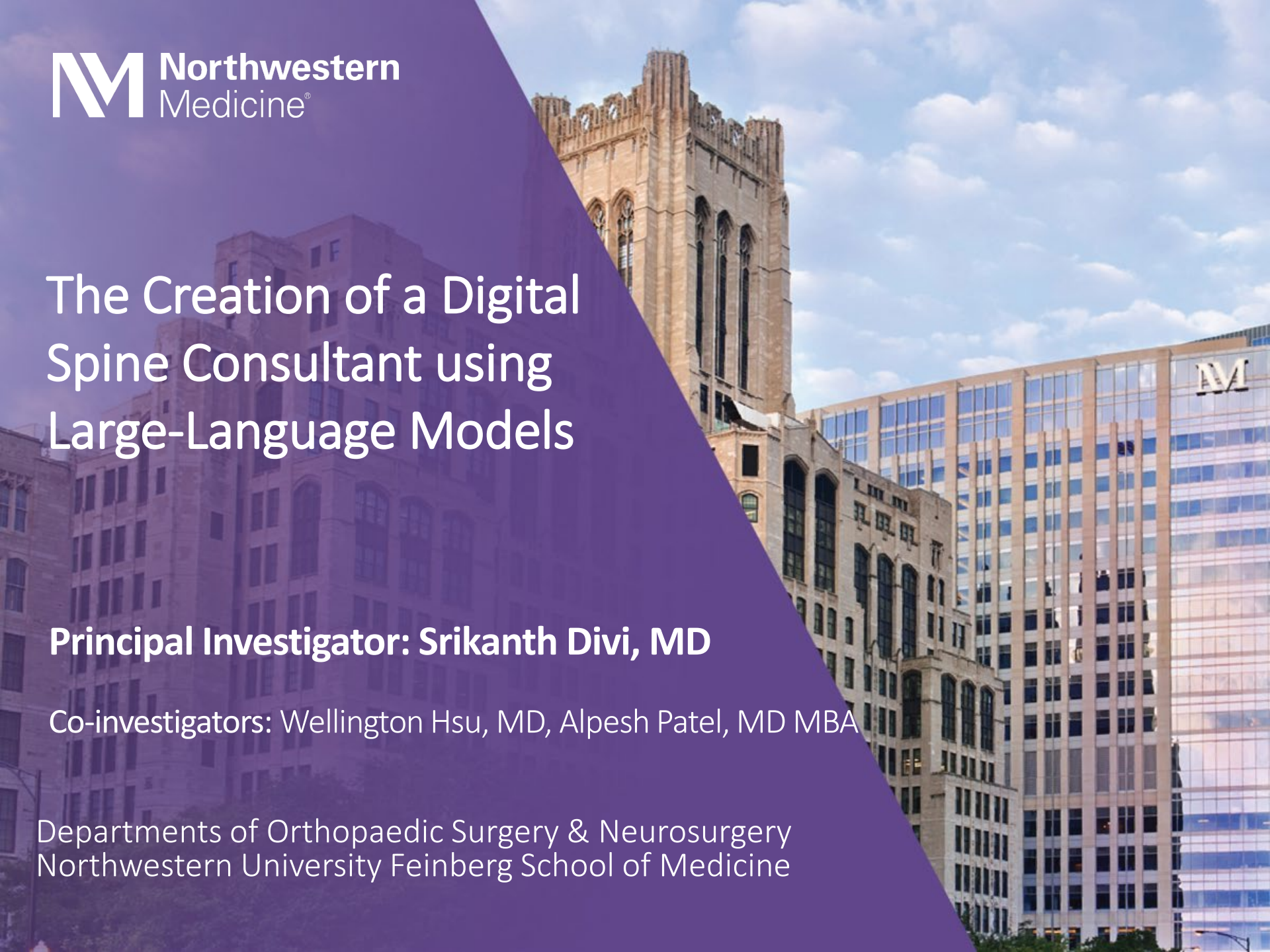


Northwestern Medicine®

The Creation of a Digital Spine Consultant using Large-Language Models

Principal Investigator: Srikanth Divi, MD

Co-investigators: Wellington Hsu, MD, Alpesh Patel, MD MBA

Departments of Orthopaedic Surgery & Neurosurgery
Northwestern University Feinberg School of Medicine

Co-authors

Cade Bennett, BA

Aditya Khandeshi, BS

Austin Chen, BS

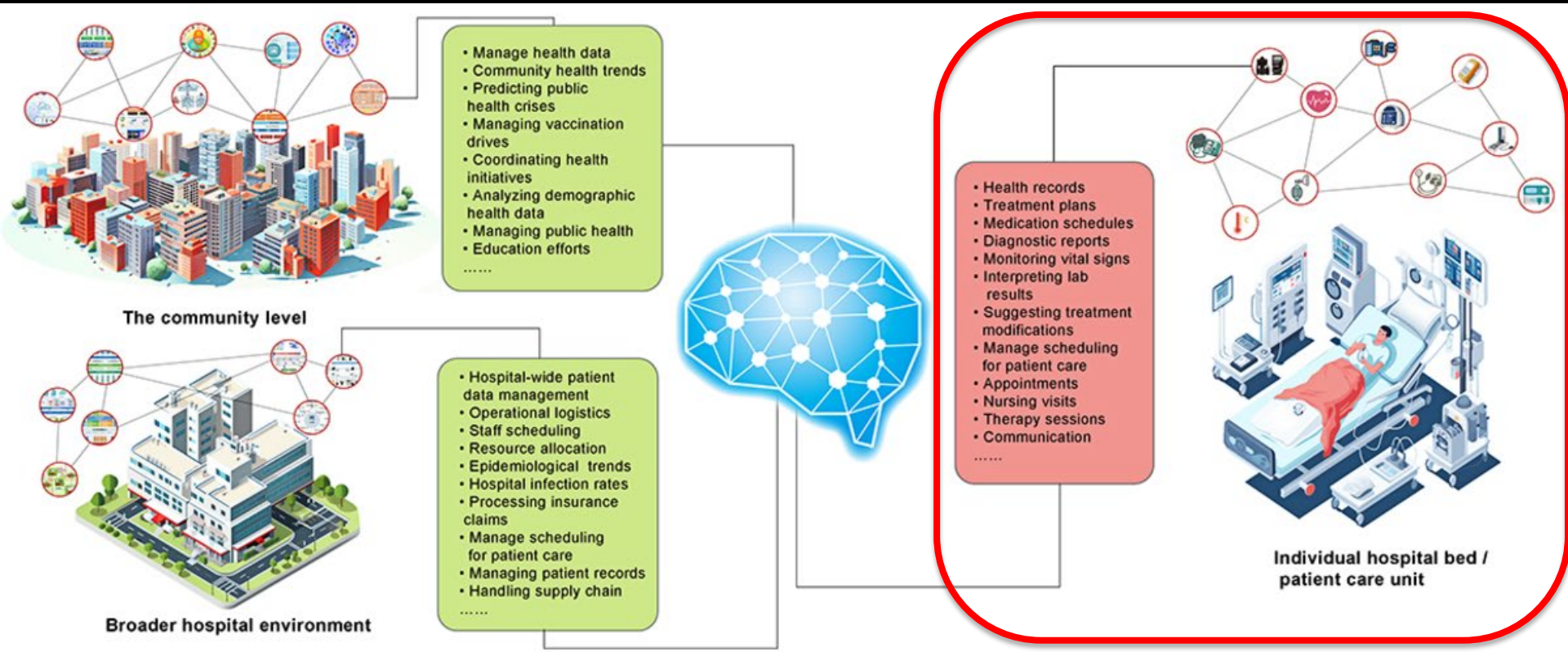
Theodore Joaquin, BA

Disclosures

Srikanth Divi	Consultant for Atec Spine, Kuros Biosciences
Co-Authors	Disclosures in Program

Background

Clinical Context



Zhang K et al. *J Med Internet Res. 2025*

Northwestern Medicine

NORTHWESTERN DEPARTMENT OF ORTHOPAEDIC SURGERY

How can LLMs help?

- Compensate for specialist shortage
- Triage patients effectively
- Reduce clinical burden
- Provide individualized care

- Reduce administrative burden
- Medical education
- Patient education

Scientific Aims

Aim #1:

Develop LLM to guide patient care for cervical spine consults

Aim #2:

Compare accuracy of several widely-available LLMs

Methods

Methods

Built and queried repository of resident consults (2014-2024)

Patient notes, labs, and imaging reports from +/- 1 day within consult date

Redaction of assessment/plan in notes (ChatGPT-o4-mini)

Microsoft Azure Foundry: ChatGPT-5, DeepSeek-R1-0528, Llama-4-Maverick

Models graded for Clinical Accuracy

Methods

Grading:

- Clinical summary accuracy
  - Likert scale from 1 (least) to 5 (most accurate)
- Recommendation for surgery
  - 0 (incorrect) or 1 (correct)

- Model reasoning for nonoperative care
  - 1 = adequate reasoning
  - 2 = highly nuanced reasoning

Prompt Engineering

“You are an assistant to a spine surgeon who screens patients referred for spine surgery. You are provided with all of the notes, lab reports, and imaging reports for a patient receiving consultation and evaluation for spine surgery. Each report is separated by the delimiter '----------', with the Assessment and Plan sections masked by the delimiter '<REC>'. Ignoring these Assessments, please look through each note and report and create a summary report highlighting the following:

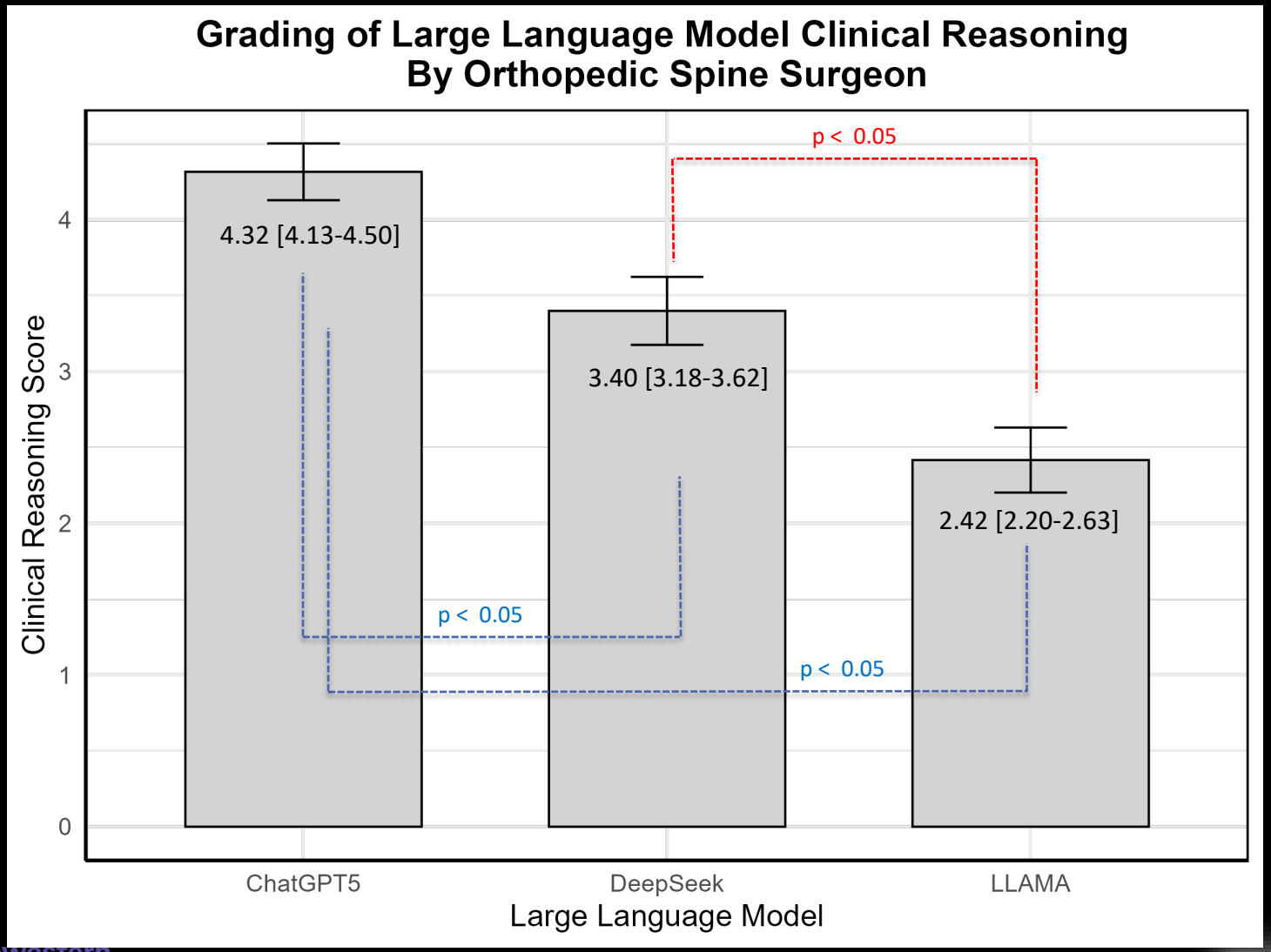
1. Determine whether the patient needs a surgical approach (i.e. spine surgery).
2. If a surgical procedure is required, provide a specific procedural recommendation for the patient.
3. Provide your rationale for points 1 and 2. This rationale can include diagnoses and interpretation but should also include medical information that synthesizes these interpretations (if used) to support the recommendations with medical evidence.”

Results

Cervical Spine Consults

Variable	Subclass	Overall
Total		60
Consult Type n (%)	Degenerative Disease	24 (41.4)
	Infection	4 (6.9)
	Oncology	8 (13.8)
	Trauma	22 (37.9)
Urgent Surgery? n (%)	Yes	8 (13.6)
	No	52 (86.4)

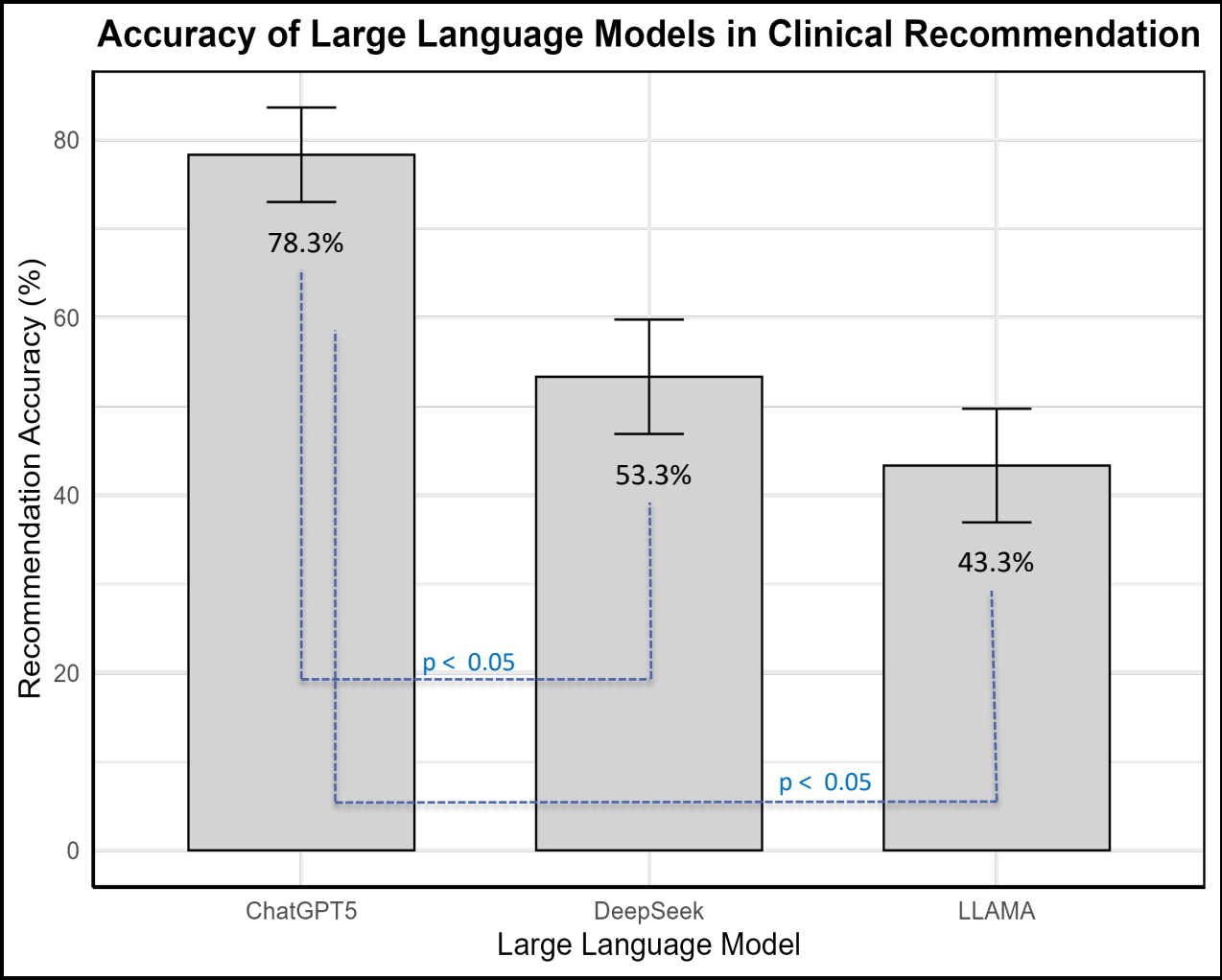
Clinical Accuracy



Northwestern Medicine

NORTHWESTERN DEPARTMENT OF ORTHOPAEDIC SURGERY

Correct Surgery Recommendation



Northwestern Medicine

NORTHWESTERN DEPARTMENT OF ORTHOPAEDIC SURGERY

Rationale for nonop treatment

	ChatGPT-5	DeepSeek-R1	Llama-4	p-Value
n (%)	41	22	15	
Adequate (1)	4 (10.5)	9 (40.9)	13 (86.7)	<0.001
Superior (2)	34 (89.5)	13 (59.1)	2 (13.3)	

Example – Resident Blurb

69M w/ PMH of chronic cervical myelopathy 2/2 DISH s/p C2-T2 PSH (2013, Dr. *****) w/ residual R sided weakness, metastatic colon adenocarcinoma s/p diverting ileostomy w/ mets to liver and lungs, prostate adenocarcinoma s/p XRT (2017), DVT/PE on apixaban, CHF, T2DM, CKD stage V, HTN presenting with acute-on-chronic lower back pain, BLE weakness, paresthesias, and recurrent falls for about 1 week. Denies saddle anesthesia, urinary incontinence. Endorses chronic BUE/BLE weakness that has progressively worsened since 2013. Denies fevers/chills. On exam: cervical incisions well-healed. Motor exam: RUE w/ 4-/5 strength throughout w/ exception of 3/5 strength w/ small finger abduction. LUE w/ 4+/5 strength throughout w/ exception of 4-/5 strength w/ small finger abduction. RLE w/ 4/5 strength throughout. LLE w/ 2/5 throughout w/ expeption of 3/5 ankle dorsiflexion and great toe dorsiflexion. Subjectively diminished sensation throughout all dermatomes in RUE, C7 distribution in LUE, and L3 distribution in RLE. +Ankle clonus bilaterally. +Hoffman's RUE. Equivocal Babinski bilaterally. Imaging demonstrates prior C2-T2 decompression and fusion, cervical myelomalacia, multilevel degenerative changes w/ notable central stenosis at T11-T12 and central stenosis at L2-L3.

No acute orthopaedic intervention-f/u as outpatient with **** to discuss non-operative management

Example – Llama-4

Patient Summary

**** is a 69-year-old male with a complex medical history that includes chronic cervical myelopathy secondary to Diffuse Idiopathic Skeletal Hyperostosis (DISH) with prior cervical spine surgery (C2-T2 posterior spinal fusion in 2013), metastatic colon adenocarcinoma with liver and lung metastases, prostate adenocarcinoma treated with radiation therapy, deep vein thrombosis/pulmonary embolism (DVT/PE) on anticoagulation therapy with apixaban, heart failure with preserved ejection fraction (HFpEF), type 2 diabetes mellitus (T2DM), chronic kidney disease (CKD) stage 5, and hypertension (HTN).

Recommendations and Rationale

1. **Surgical Approach**: Given the patient's complex medical history, including metastatic colon cancer and significant comorbidities, a surgical intervention for his spinal stenosis would be considered high-risk. However, the progression of L2-L3 spinal canal stenosis and the patient's symptoms may necessitate a surgical evaluation for potential decompression.

2. **Specific Procedural Recommendation**: A lumbar decompression surgery at L2-L3 could be considered to alleviate the patient's lower extremity weakness and back pain, provided that his overall medical condition allows for the risks associated with surgery.

3. **Rationale**: The rationale for considering surgery is based on the progression of spinal canal stenosis at L2-L3, as evidenced by recent MRI findings, and the patient's clinical presentation of increasing lower extremity weakness and back pain. The decision to proceed with surgery must be weighed against the patient's significant comorbidities, including his metastatic cancer, CKD stage 5, and cardiovascular disease.

Example – DeepSeek-R1

1. **Need for Surgical Approach**
No, spine surgery is not currently indicated.

2. **Procedural Recommendation**
N/A (No surgical intervention recommended).

3. **Rationale**
Key Medical Evidence Synthesis:
- **Stable Spinal Pathology**:
  - **Imaging (MRI, 03/13/2023)**:
    - **Cervical**: Stable myelomalacia and prior C2-T2 fusion without new cord compression.
    - **Thoracic (T11-T12)**: Moderate-severe stenosis with cord signal changes (edema/myelomalacia), **unchanged** from prior studies (02/09/2023).
    - **Lumbar (L2-L3)**: Moderate-severe spinal canal stenosis due to disc extrusion and facet degeneration, **stable** since February 2023.
  - **No acute structural progression** warranting intervention.

- **Neurological Findings**:
  - **Motor/Sensory Deficits**: Asymmetric weakness (R > L) in extremities, diminished sensation in RUE/LLE. **Unchanged** from baseline (residual from 2013 cervical surgery).
  - **No New Deficits**: Acute left-hand weakness (03/12) resolved after discontinuing baclofen, confirming **non-structural etiology** (likely medication toxicity).
  - **Absence of Red Flags**: Denies saddle anesthesia, urinary/bowel incontinence, or progressive myelopathy.

Conclusion:
The absence of acute neurological deterioration, stable imaging findings, high surgical risk, and active oncologic management support **non-operative management**. Focus should remain on medical optimization, chemotherapy, and rehabilitation.

Example – ChatGPT-5

Summary for Spine Surgery Triage

1) Does this patient need spine surgery now?
- No urgent or immediate spine surgery is indicated at this time. Elective surgery is not recommended currently given the patient multi-level, largely chronic disease, absence of an acute compressive syndrome, and very high perioperative risk from multiple comorbidities.

2) If surgery becomes necessary, what specific procedure would be appropriate?
- If focal decompression is pursued for persistent, disabling symptoms or progressive objective deficits despite maximal nonoperative care:
  Primary target: L2-L3 decompressive laminectomy with resection of hypertrophied ligamentum flavum and epidural lipomatosis, with removal of the central disc extrusion. Fusion would not be routinely planned unless intraoperative instability is encountered (current imaging describes normal alignment without spondylolisthesis).
  Secondary consideration (only if clear progression of thoracic myelopathy attributable to this level): T11-T12 posterior decompression (laminectomy with medial facetectomies; resection of the right-sided synovial cyst/ligamentous thickening causing dorsolateral cord encroachment). This carries substantially higher neurologic risk and should be reserved for demonstrated progression.

3) Rationale
Why no urgent surgery now
- No acute cauda equina or rapidly progressive myelopathy on serial exams or imaging. The thoracic cord signal abnormality at T11-T12 is chronic and stable; L2-L3 stenosis progressed between November 2022 and February 2023 but is unchanged since February and March 2023.
- The recent transient worsening in upper extremity function and mentation correlated with baclofen toxicity and improved off baclofen.
- Pain is being controlled medically; the patient remains significantly deconditioned but participates in rehabilitation when medically optimized.

If surgery is reconsidered
- Ensure goals-of-care alignment and that anticipated neurologic/functional benefit justifies risk.
- Optimize medical comorbidities: treat infections; improve anemia if possible; control blood pressure; coordinate with cardiology/renal as needed.
- Perioperative anticoagulation plan for apixaban with attention to CKD; thromboembolic history makes bridging decisions complex.
- Start with the least extensive, most symptomatic level L2-3 decompression with removal of epidural fat and ligamentum flavum) and avoid fusion unless instability is encountered.
- Reserve thoracic decompression at T11-12 for demonstrated progression of myelopathic signs attributable to that level.

Discussion

General Performance

Clinical Accuracy:

- ChatGPT-5 > DeepSeek > Llama

Correct Surgery Recommendation

- ChatGPT-5 > DeepSeek = Llama

Nuanced Clinical Reasoning

- ChatGPT-5 >> DeepSeek

Highlights

- Models succinctly organized comorbidities and risk factors
- Longitudinal view of patient
- ChatGPT-5 routinely suggested next steps for patients
  - Imaging
  - Timeline for followup
  - Symptoms/signs to monitor

Pitfalls

- Models underperformed for myelopathy or cord compression
  - Quick to suggest operative management
  - Perhaps heavily rely on radiology interpretation
- Also underperformed for geriatric trauma (C2 fractures)
  - Suggested operative management for majority of geriatric odontoid fractures
- Potentially modifiable
  - Institutional protocols
  - Surgeon specific care pathways

Conclusion

- Higher end models performed with high accuracy
- Provided excellent clinical rationale
- Demonstrates probable satisfactory use as a clinical assistant

Next Steps

- Incorporate Claude, possible Google Gemini 3
- Engineer different prompts
- Multiple reviewers
- Analyze all cervical spine patients

Thank You!

NORTHWESTERN
DEPARTMENT OF ORTHOPAEDIC SURGERY