RQ3b: Domain-Specific Performance

Domain	Test AUROC (Full Model)
math	0.797
is_agents	0.703
psychology	0.671
finance	0.632
medicine	0.619

Mean: 0.685