Cross-phenomenon overlap in BLiMP

Figure showing horizontal bar chart with BLiMP Category on the y-axis and Percentage of units on the x-axis (0 to 100).

Categories (top to bottom): DET-N Agreement, S-V Agreement, Anaphor Agreement, Ellipsis, Filler–Gap, Control/Raising, Argument Structure, NPI Licensing, Island Effects, Binding, Quantifiers, Irregular Forms.

Values:
- DET-N Agreement: 68.91% / 10.10%
- S-V Agreement: 40.29% / 9.47%
- Anaphor Agreement: 23.10% / 7.78%
- Ellipsis: 13.64% / 4.50%
- Filler–Gap: 13.18% / 5.16%
- Control/Raising: 13.14% / 6.76%
- Argument Structure: 12.61% / 7.28%
- NPI Licensing: 6.78% / 3.45%
- Island Effects: 4.91% / 2.26%
- Binding: 6.54% / 5.22%
- Quantifiers: 2.14% / 3.02%
- Irregular Forms: 4.32% / 5.65%

Models:
- Falcon3-3B-Base (red circle)
- Llama-3.2-3B (green square)
- Mistral-7B-v0.3 (purple triangle)
- Phi-4-mini-instruct (pink diamond)
- deepseek-llm-7b-base (yellow inverted triangle)
- gemma-3-4b-pt (brown right triangle)
- gpt2-xl (gray left triangle)

Averages:
- Overlap within category (blue)
- Overlap with other categories (orange)