Experimental Settings 1
Arepair Benchmark
(38 Specs)

Arepair
9 cases (24%)

BeAFix
24 cases (63%)

ATR
14 cases
(37%)

ICEBAR
21 cases (55%)

chatGPT
27 cases (71%)
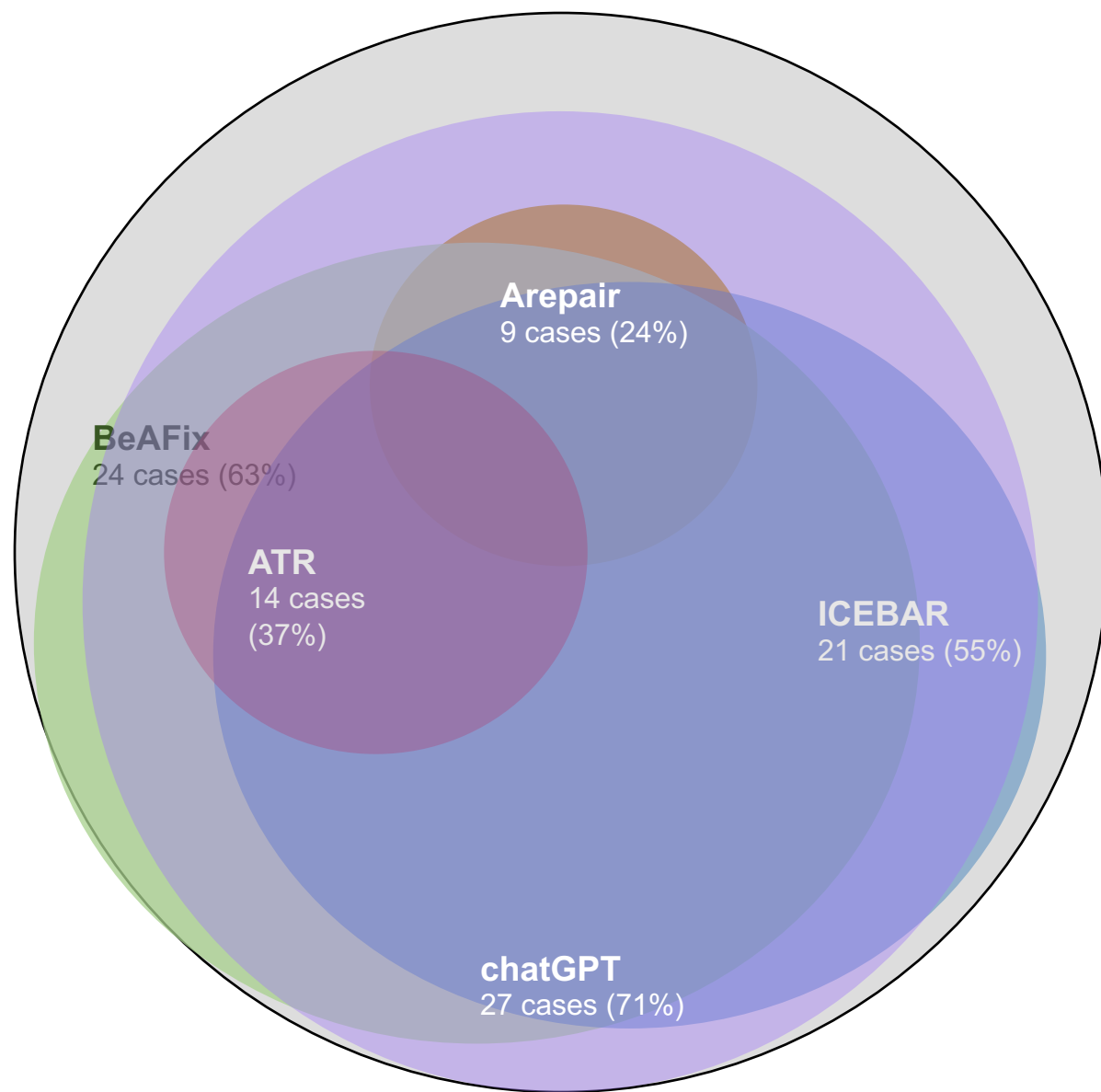
Experimental Settings 2
Arepair Benchmark
(38 Specs)

Arepair
9 cases (24%)

BeAFix
24 cases
(63%)

ATR
14 cases
(37%)

chatGPT
19 cases (50%)

ICEBAR
21 cases (55%)
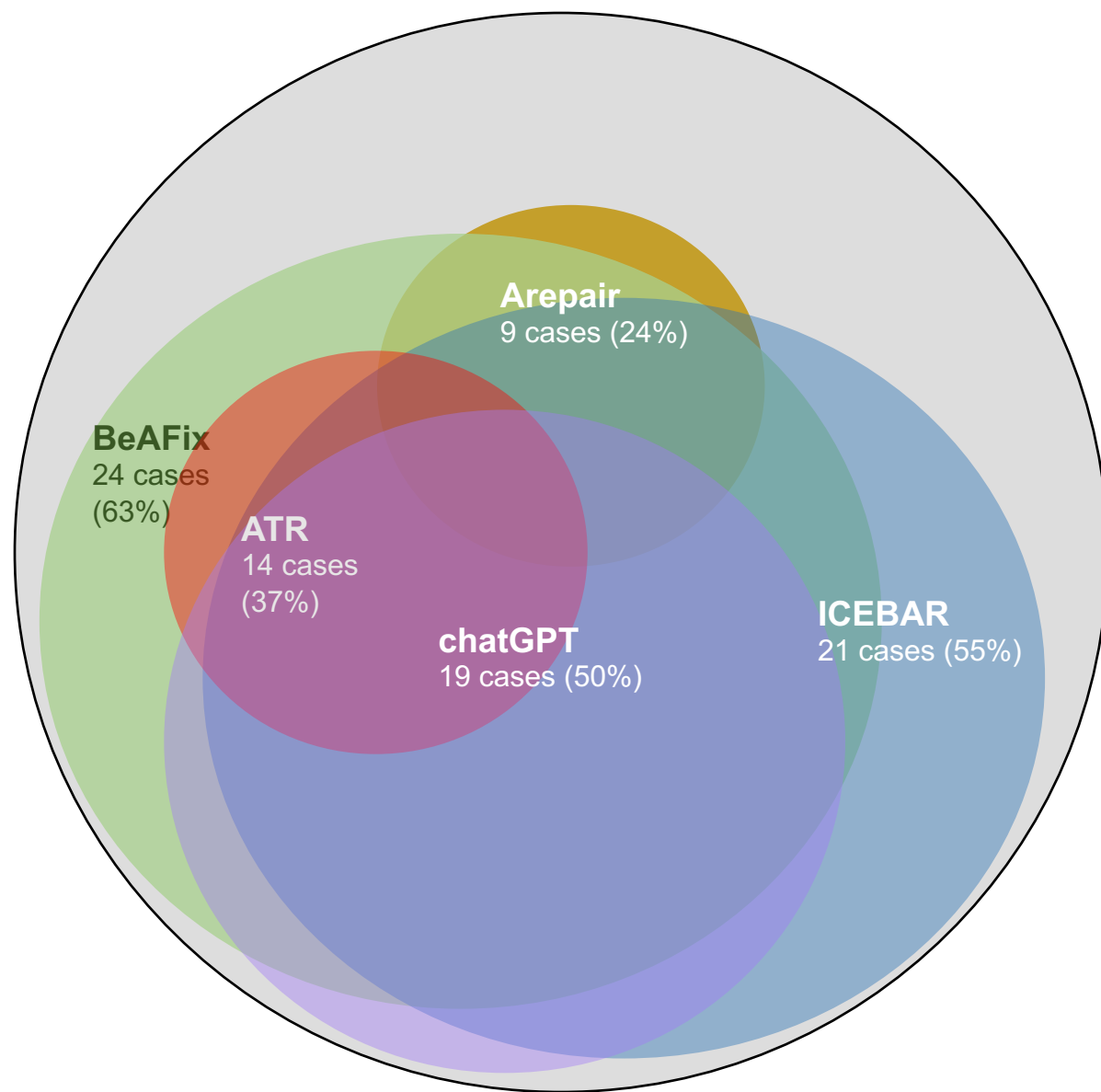
Experimental Settings 3
Arepair Benchmark
(38 Specs)

Arepair
9 cases (24%)

BeAFix
24 cases
(63%)

ATR
14 cases
(37%)

ICEBAR
21 cases (55%)

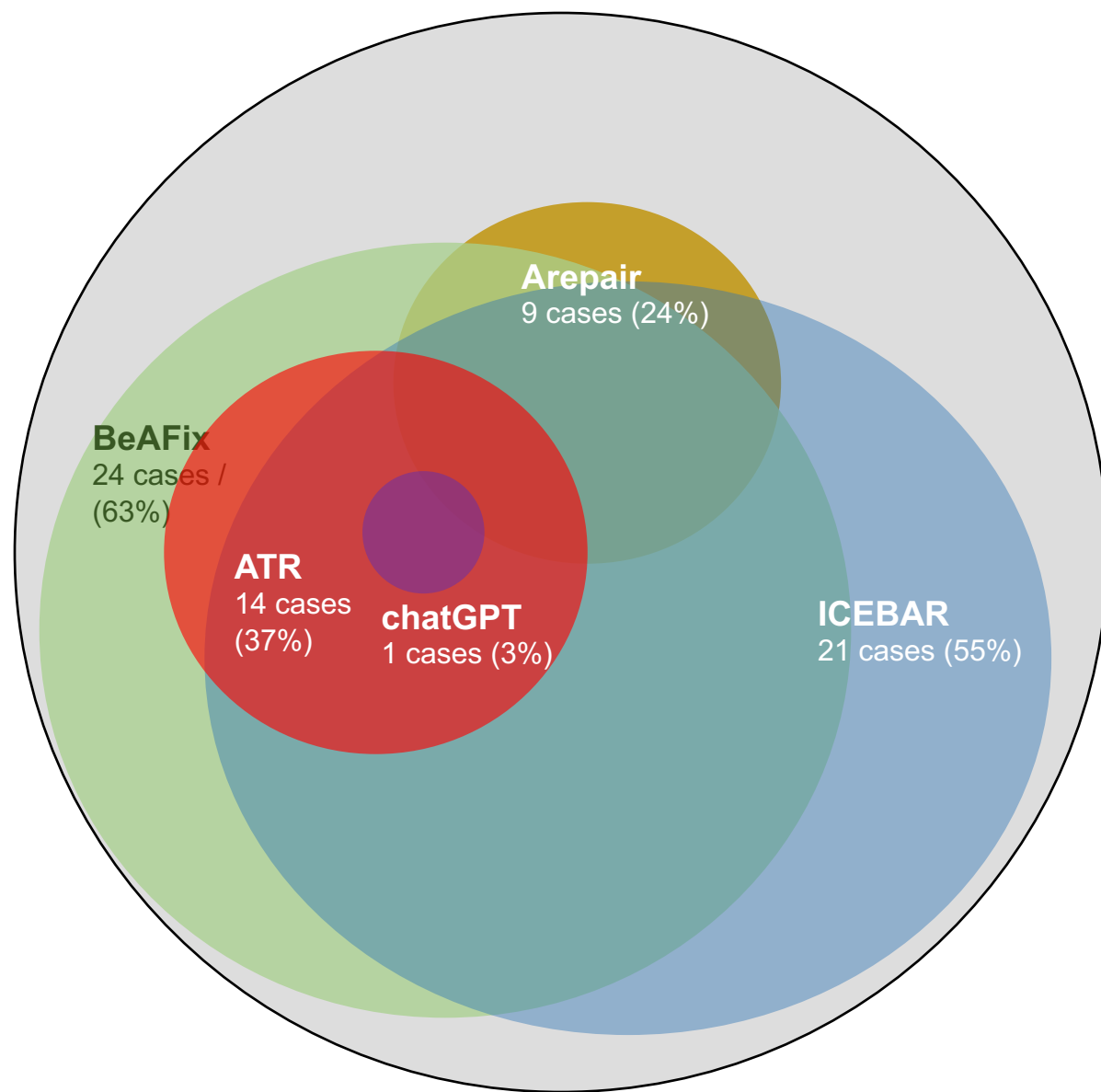chatGPT
23 cases (60%)
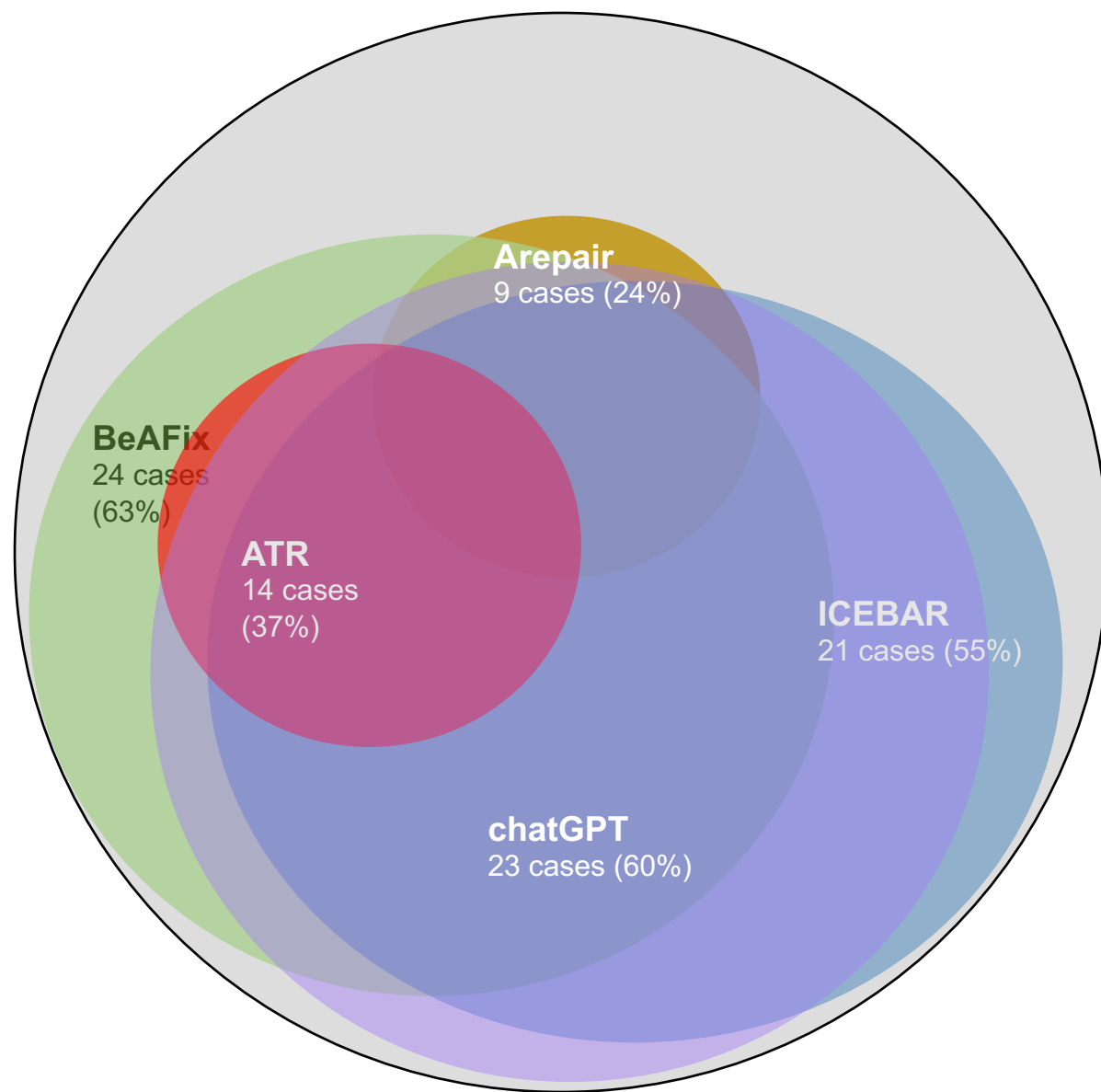
Experimental Settings 4
Arepair Benchmark
(38 Specs)

Arepair
9 cases (24%)

BeAFix
24 cases /
(63%)

ATR
14 cases
(37%)

chatGPT
1 cases (3%)

ICEBAR
21 cases (55%)

Experimental Settings 5
Arepair Benchmark
(38 Specs)

Arepair
9 cases (24%)

BeAFix
24 cases
(63%)

ATR
14 cases
(37%)

chatGPT
5 cases (13%)

ICEBAR
21 cases (55%)