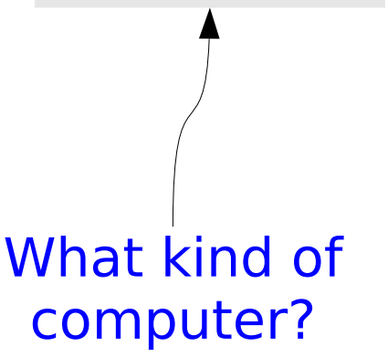
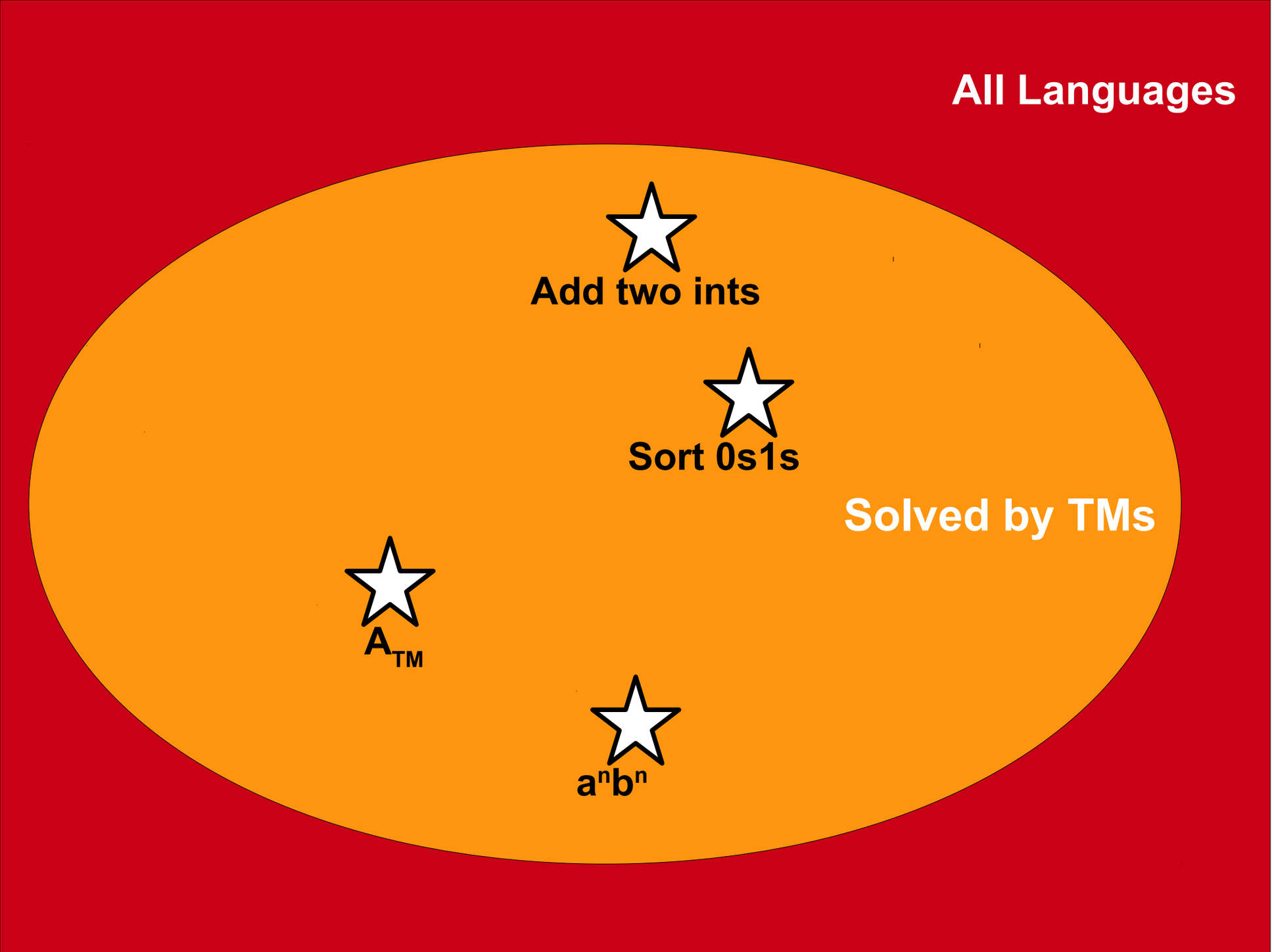
Turing Machines

Part Three

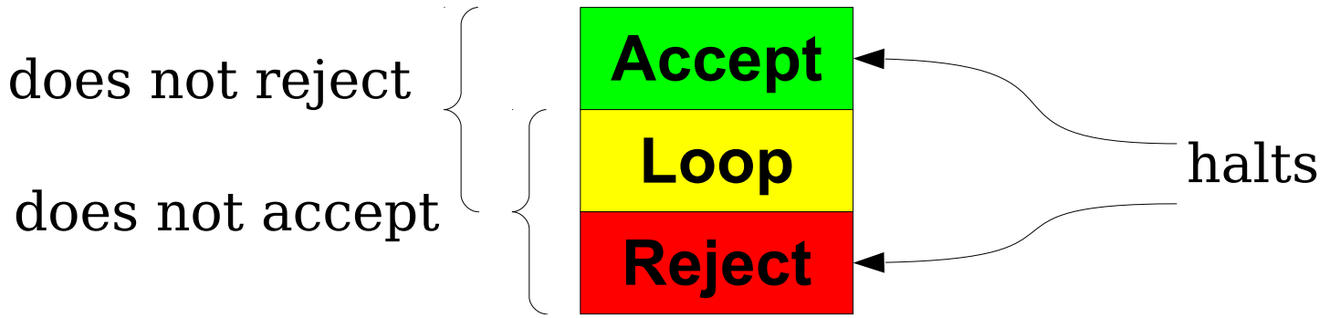
What problems can we solve with a computer?




All Languages
Add two ints
Sort 0s1s
Solved by TMs
A_{TM}
$a^n b^n$

Possible Turing Machine Outcomes

- Let M be a Turing machine.
- M ***accepts*** a string w if it enters an accept state when run on w.
- M ***rejects*** a string w if it enters a reject state when run on w.
- M ***loops infinitely*** (or just ***loops***) on a string w if when run on w it enters neither an accept nor a reject state. (such a w is **not** in the language of this TM)



Very important terminology:

Recognizable Languages (**RE**)

- A language is called ***recognizable*** if it is the language of some TM.
  - For any $w \in \mathscr{L}(M)$, M accepts w.
  - For any $w \notin \mathscr{L}(M)$, M does not accept w.
    - M **might reject**, or it **might loop forever**.

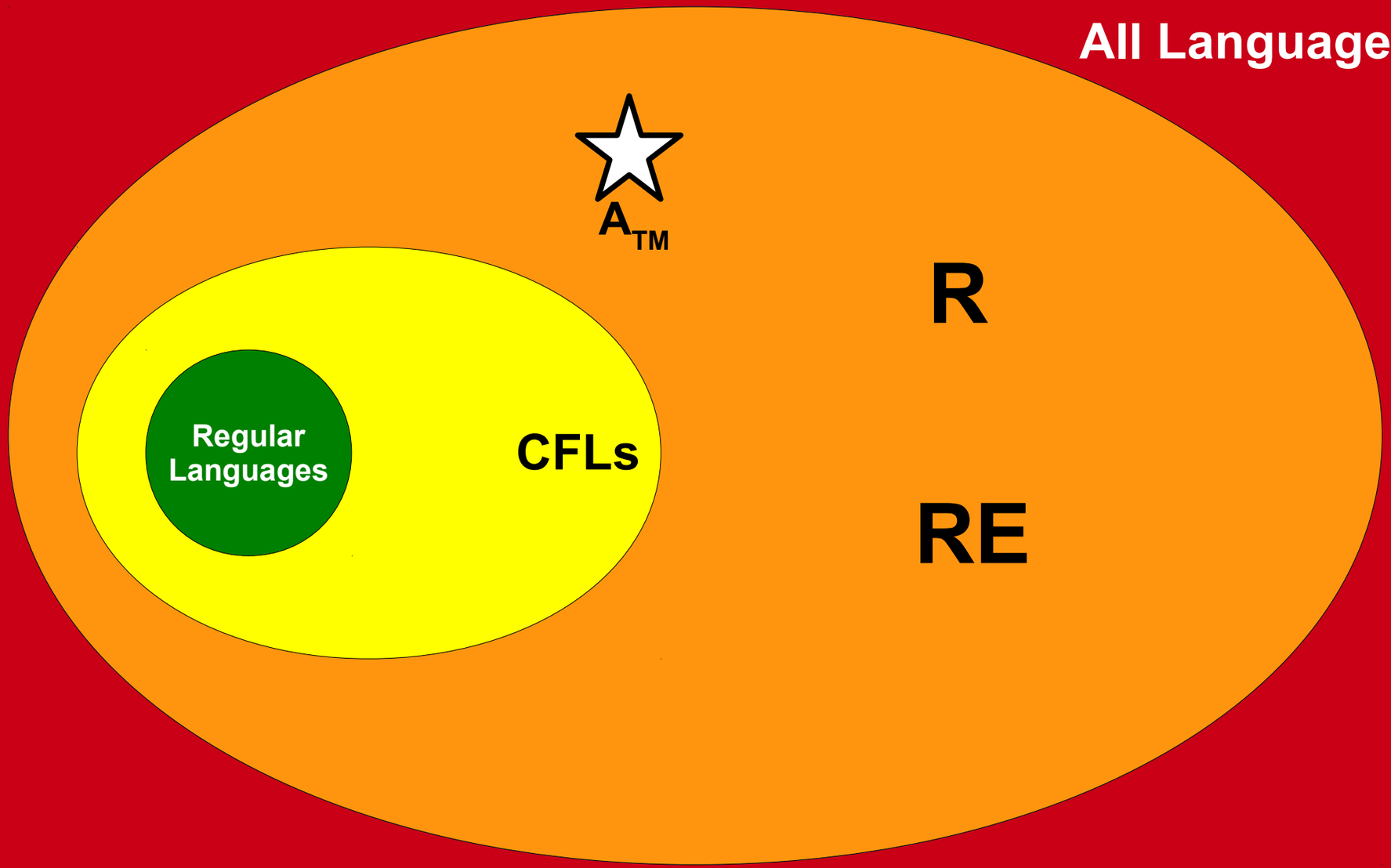
Decidable Languages (**R**)

- A language L is called ***decidable*** if there exists a decider M such that $\mathscr{L}(M) = L$.
    - Decider machines are implemented in a way that they have no danger/possibility of looping forever.

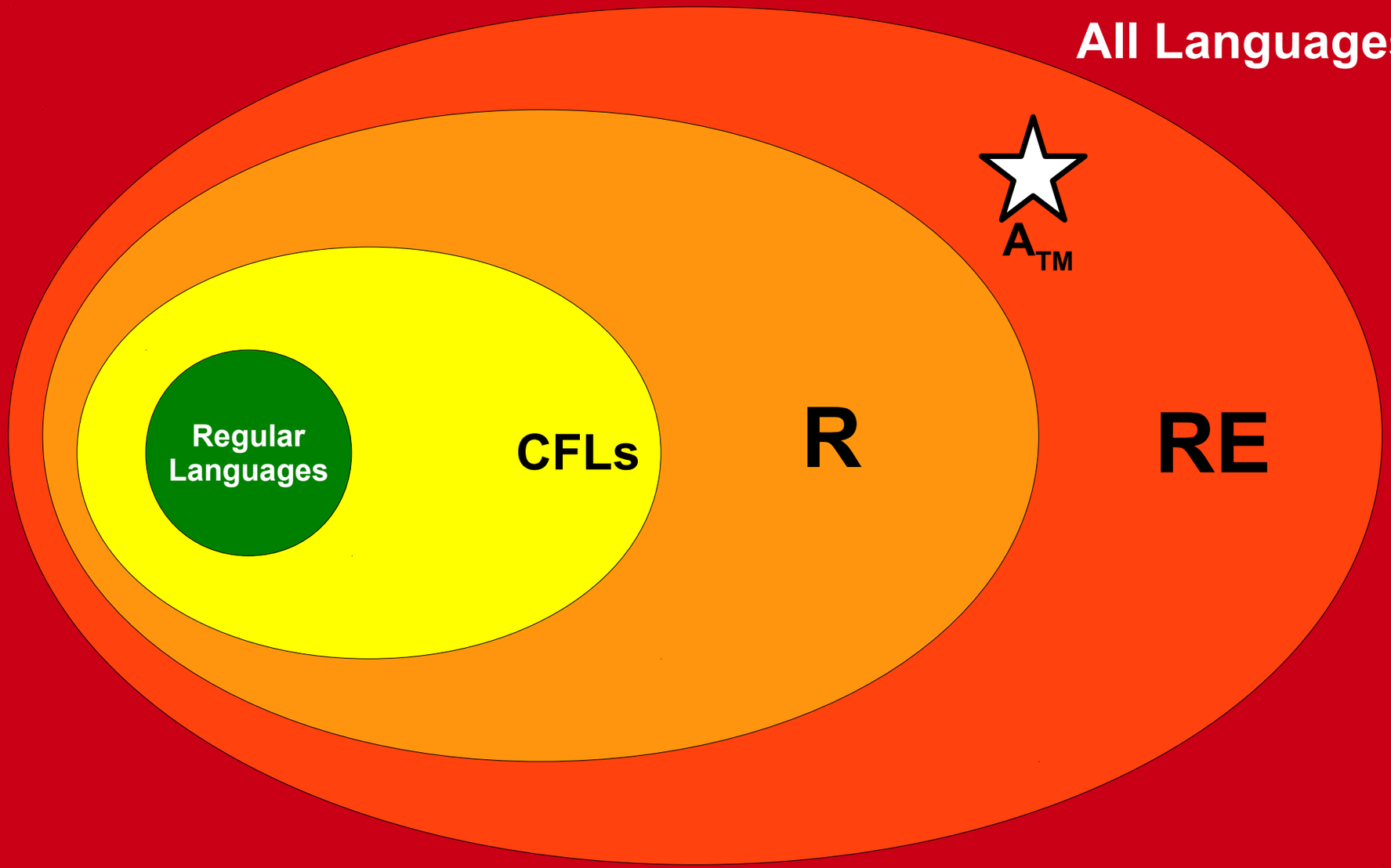
R and **RE** Languages

- Every decider is a Turing machine, but not every Turing machine is a decider.
- This means that **R** ⊆ **RE**.
- But is it a *strict* subset?
- That is, if you can just confirm “yes” answers to a problem, can you necessarily *solve* that problem?

Which Picture is Correct?



Which Picture is Correct?



Self-Referential :Danger:

- Remember our self-referential code from Monday? (EqualsMe, AmIEven, etc)

- So, I hope we've convinced you that there's nothing magic, impossible, or scary about a program getting a string version of its own code.

- But, there are some dragons in the land of self-referential things….

True or false:

```
"This string is 34 characters
            long."
```

True or false:

```
"This string is 34 characters long."
 12345678901234567890123456789012345
```

True or false:

"This sentence is written in blue."

True or false:

"This sentence is false."

Happy Story Time

In a certain isolated town, every house has a lawn and the city requires them all to be mowed. The town has only one gardener, who is also a resident of the town, and this gardener mows the lawns of residents iff they do not mow their own lawn.

Happy Story Time

In a certain isolated town, every house has a lawn and the city requires them all to be mowed. The town has only one gardener, who is also a resident of the town, and this gardener mows the lawns of residents iff they do not mow their own lawn.

True or false: The gardener mows their own lawn.


MY NOSE WILL
GROW NOW!

Self-Reference in Set Theory

- Now that we know that self-reference is dangerous (i.e., can lead to paradoxes) in propositions (e.g.,“This sentence is false”), we can look at it in other domains we’ve studied, like Set Theory.

Self-Reference in Set Theory

- Now that we know that self-reference is dangerous (i.e., can lead to paradoxes) in propositions (e.g.,“This sentence is false”), we can look at it in other domains we’ve studied, like Set Theory.
  - We know that sets can contain other sets.

Self-Reference in Set Theory

- Now that we know that self-reference is dangerous (i.e., can lead to paradoxes) in propositions (e.g.,“This sentence is false”), we can look at it in other domains we’ve studied, like Set Theory.
  - We know that sets can contain other sets.
  - We never said they can’t contain themselves (“The set of all sets” and “The set of all sets with infinite cardinality” would be examples of sets that would contain themselves.)

Self-Reference in Set Theory

- Now that we know that self-reference is dangerous (i.e., can lead to paradoxes) in propositions (e.g.,“This sentence is false”), we can look at it in other domains we’ve studied, like Set Theory.
  - We know that sets can contain other sets.
  - We never said they can’t contain themselves (“The set of all sets” and “The set of all sets with infinite cardinality” would be examples of sets that would contain themselves.)
  - Since we know that self-reference is dangerous, we might want to make a set called SAFE_LIST that is the set of all sets that do *not* contain themselves, since any set on that list is safe from paradoxes.

Self-Reference in Set Theory

- Now that we know that self-reference is dangerous (i.e., can lead to paradoxes) in propositions (e.g.,“This sentence is false”), we can look at it in other domains we’ve studied, like Set Theory.
  - We know that sets can contain other sets.
  - We never said they can’t contain themselves (“The set of all sets” and “The set of all sets with infinite cardinality” would be examples of sets that would contain themselves.)
  - Since we know that self-reference is dangerous, we might want to make a set called SAFE_LIST that is the set of all sets that do *not* contain themselves, since any set on that list is safe from paradoxes.

True or False: SAFE_LIST does *not* contain itself.

Self-Defeating Objects

Proofs by Contradiction in Number Theory

- One way to think about proofs by contradiction is that they lead to a kind of “impossible” situation that is similar to the paradoxes.
- One form of proof by contradiction makes use of what is called the “self-defeating object,” a thing that we assume for the sake of contradiction exists, but we will show can’t exist because its existence contradicts itself.
  - Similar to how SAFE_LIST, or the lawn mower, contradicted their own existences.

Proofs by Contradiction in Number Theory

- Here is a simple example:
  - **Thm.** There is no greatest integer.
  - **Proof, by contradiction.** Assume for the sake of contradiction that there is a greatest integer, call it g.
  - *[Now we will use g to write a mathematical expression that is a syntactically valid mathematical expression that should be fine to write,* ***if*** **g were real***.]*
  - Let $x = g + 1$.
  - We see that $x > g$.
  - But this is a contradiction, because g is the greatest integer.
  - So the assumption is false and the theorem is true. ■

Proofs by Contradiction in Number Theory

- Here is a si
  - **Thm.** The
  - **Proof, by contradiction.** Assume for the sake of contradiction that there is a greatest integer, call it g.
  - *[Now we will use g to write a mathematical expression that is a syntactically valid mathematical expression that should be fine to write,* ***if*** **g were real***.]*
  - **Let $x = g - 1$.**
  - We see that $x < g$.
  - There is no contradiction, so g actually is the greatest integer!
  - So the theorem is false, and there is a greatest integer. ■

What is the problem with this (modified) proof?

Proofs by Contradiction in Number Theory

- Here is a simple ex
  - **Thm.** There is no g
  - **Proof, by contradi**
    contradiction that t
  - *[Now we will use g*
    that is a syntactical
    that should be fine t
  - **Let $x = g + 1$.**
  - We see that $x > g$.
  - But this is a contradiction, because g is the greatest integer.
  - So the assumption is false and the theorem is true. ■

Observation: "Fixing" the proof by changing the math from addition to subtraction doesn't actually fix anything, because it wasn't the math that was the problem. It was g itself.

We chose addition on purpose because it's what we needed to do to **expose the existing problem with *g*.**

More Self-Reference!

(this time with Turing Machines)

A Decider for A_{TM}?

- ***Recall:*** A_{TM} is the language of the universal Turing machine.
- We know that $\langle M, w \rangle \in A_{TM}$ if and only if M accepts w.
- The universal Turing machine U_{TM} is a *recognizer* for A_{TM}. Could we build a *decider* for A_{TM}?

A Decider for A_{TM}?

- Suppose that $A_{TM} \in \mathbf{R}$.
- Formally, this means that there is a TM that decides A_{TM}.
- Intuitively, this means that there is a TM that takes as input ⟨M, w⟩, then
  - accepts if *M* accepts *w*, and
  - rejects if *M* does not accept *w*.
  - (i.e., infinite looping is not possible)

A Decider for A_{TM}: `willAccept`

- To make the previous discussion more concrete, let's talk about this hypothetical decider for A_{TM} as a computer program
- If A_{TM} is decidable, we could construct a function

```
bool willAccept(string program /*M*/,
                string input   /*w*/)
```

  that returns true if the program will accept the input and false otherwise (*never infinite looping*).
- **Hypothetically,** if `willAccept` existed, what could we do with it?

If A_{TM} is decidable, we could construct a function

```
bool willAccept(string program /*M*/,
                string input   /*w*/)
```

that returns true if the program will accept the input and false otherwise (*never infinite looping*).

How many of the following statements are true?

`willAccept("main() {   accept();   }",   "Emu")` returns true.
`willAccept("main() {   reject();   }",   "Yak")` returns false.
`willAccept("main() { while (true) {} }", "Cow")` loops forever.
`willAccept("main() { while (true) {} }", "Cow")` returns true.
`willAccept("main() { while (true) {} }", "Cow")` returns false.

What does this program do?

```
bool mystery(string input) {
   string me = mySource();
   if (willAccept(me, input)) {
      return false;
   } else {
      return true;
   }
}
```

If A_{TM} is decidable, we could construct a function

```
bool willAccept(string program /*M*/,
                string input   /*w*/)
```

that returns true if the program will accept the input and false otherwise (*never infinite looping*).

What does this program do?

```
bool mystery(string input) {
   string me = mySource();
   if (willAccept(me, input)) {
      return false;
   } else {
      return true;
   }
}
```

How many of

This prog
This prog
This prog

Try running this program on any input.
What happens if

... this program accepts its input?
It rejects the input!

... this program doesn't accept its input?
It accepts the input!

If A_{TM} is decidable, we coul

```
bool willAccept(st
                st
```

that returns true if the program will accept the input and false otherwise (*never infinite looping*).

What does this program do?

```
bool mystery(string input) {
   string me = mySource();
   if (willAccept(me, input)) {
      return false;
   } else {
      return true;
   }
}
```

Try running this program on any input.
What happens if

... this program accepts its input?
It rejects the input!

... this program doesn't accept its input?
It accepts the input!

If A_{TM} is decidable, we could construct a function

```
bool willAccept(string program /*M*/,
                string input   /*w*/)
```

that returns true if the program will accept the input and false otherwise (*never infinite looping*).

Knowing the Future

- This TM is analogous to a classical philosophical/logical paradox:

If you know what you are fated to do, can you avoid your fate?

- If A_{TM} is decidable, we can construct a TM that determines what it's going to do in the future (whether it will accept its input), then actively chooses to do the opposite.
- This leads to an impossible situation with only one resolution: **A_{TM} *must not be decidable!***

Next: writing this up as a proof

Theorem: $A_{TM} \notin \mathbf{R}$. *(Recall:* **R** *is the name of the set of decidable languages)*

Theorem: $A_{TM} \notin \mathbf{R}$. *(Recall:* **R** *is the name of the set of decidable languages)*

Proof: By contradiction; assume that $A_{TM} \in \mathbf{R}$.

Theorem: $A_{TM} \notin \mathbf{R}$.

Proof: By contradiction; assume that $A_{TM} \in \mathbf{R}$. Then there is some decider D for A_{TM}, which we can represent in software as a method `willAccept` that takes as input the source code of a program and an input, then returns true if the program accepts the input and false otherwise.

Theorem: $A_{TM} \notin \mathbf{R}$.

Proof: By contradiction; assume that $A_{TM} \in \mathbf{R}$. Then there is some decider D for A_{TM}, which we can represent in software as a method `willAccept` that takes as input the source code of a program and an input, then returns true if the program accepts the input and false otherwise.

Given this, we could then construct this program P:

```
bool mystery(string input) {
    string me = mySource();

    if (willAccept(me, input)) return false;
    else return true;
}
```

Theorem: $A_{TM} \notin \mathbf{R}$.

Proof: By contradiction; assume that $A_{TM} \in \mathbf{R}$. Then there is some decider D for A_{TM}, which we can represent in software as a method `willAccept` that takes as input the source code of a program and an input, then returns true if the program accepts the input and false otherwise.

Given this, we could then construct this program P:

```
bool mystery(string input) {
    string me = mySource();

    if (willAccept(me, input)) return false;
    else return true;
}
```

Pick an arbitrary string w and trace through the execution of program P on input w.

Theorem: $A_{TM} \notin \mathbf{R}$.

Proof: By contradiction; assume that $A_{TM} \in \mathbf{R}$. Then there is some decider D for A_{TM}, which we can represent in software as a method `willAccept` that takes as input the source code of a program and an input, then returns true if the program accepts the input and false otherwise.

Given this, we could then construct this program P:

```
bool mystery(string input) {
    string me = mySource();

    if (willAccept(me, input)) return false;
    else return true;
}
```

Pick an arbitrary string w and trace through the execution of program P on input w. If `willAccept(me, input)` returns true, then P must accept its input w.

Theorem: $A_{TM} \notin \mathbf{R}$.

Proof: By contradiction; assume that $A_{TM} \in \mathbf{R}$. Then there is some decider D for A_{TM}, which we can represent in software as a method `willAccept` that takes as input the source code of a program and an input, then returns true if the program accepts the input and false otherwise.

Given this, we could then construct this program P:

```
bool mystery(string input) {
    string me = mySource();

    if (willAccept(me, input)) return false;
    else return true;
}
```

Pick an arbitrary string w and trace through the execution of program P on input w. If `willAccept(me, input)` returns true, then P must accept its input w. However, in this case P proceeds to reject its input w. Otherwise, if `willAccept(me, input)` returns false, then P must not accept its input w. However, in this case P proceeds to accept its input w.

In both cases we reach a contradiction, so our assumption must have been wrong.

Theorem: $A_{TM} \notin \mathbf{R}$.

Proof: By contradiction; assume that $A_{TM} \in \mathbf{R}$. Then there is some decider D for A_{TM}, which we can represent in software as a method `willAccept` that takes as input the source code of a program and an input, then returns true if the program accepts the input and false otherwise.

Given this, we could then construct this program P:

```
bool mystery(string input) {
    string me = mySource();

    if (willAccept(me, input)) return false;
    else return true;
}
```

Pick an arbitrary string w and trace through the execution of program P on input w. If `willAccept(me, input)` returns true, then P must accept its input w. However, in this case P proceeds to reject its input w. Otherwise, if `willAccept(me, input)` returns false, then P must not accept its input w. However, in this case P proceeds to accept its input w.

In both cases we reach a contradiction, so our assumption must have been wrong. Therefore, $A_{TM} \notin \mathbf{R}$.

Theorem: $A_{TM} \notin \mathbf{R}$.

Proof: By contradiction; assume that $A_{TM} \in \mathbf{R}$. Then there is some decider D for A_{TM}, which we can represent in software as a method `willAccept` that takes as input the source code of a program and an input, then returns true if the program accepts the input and false otherwise.

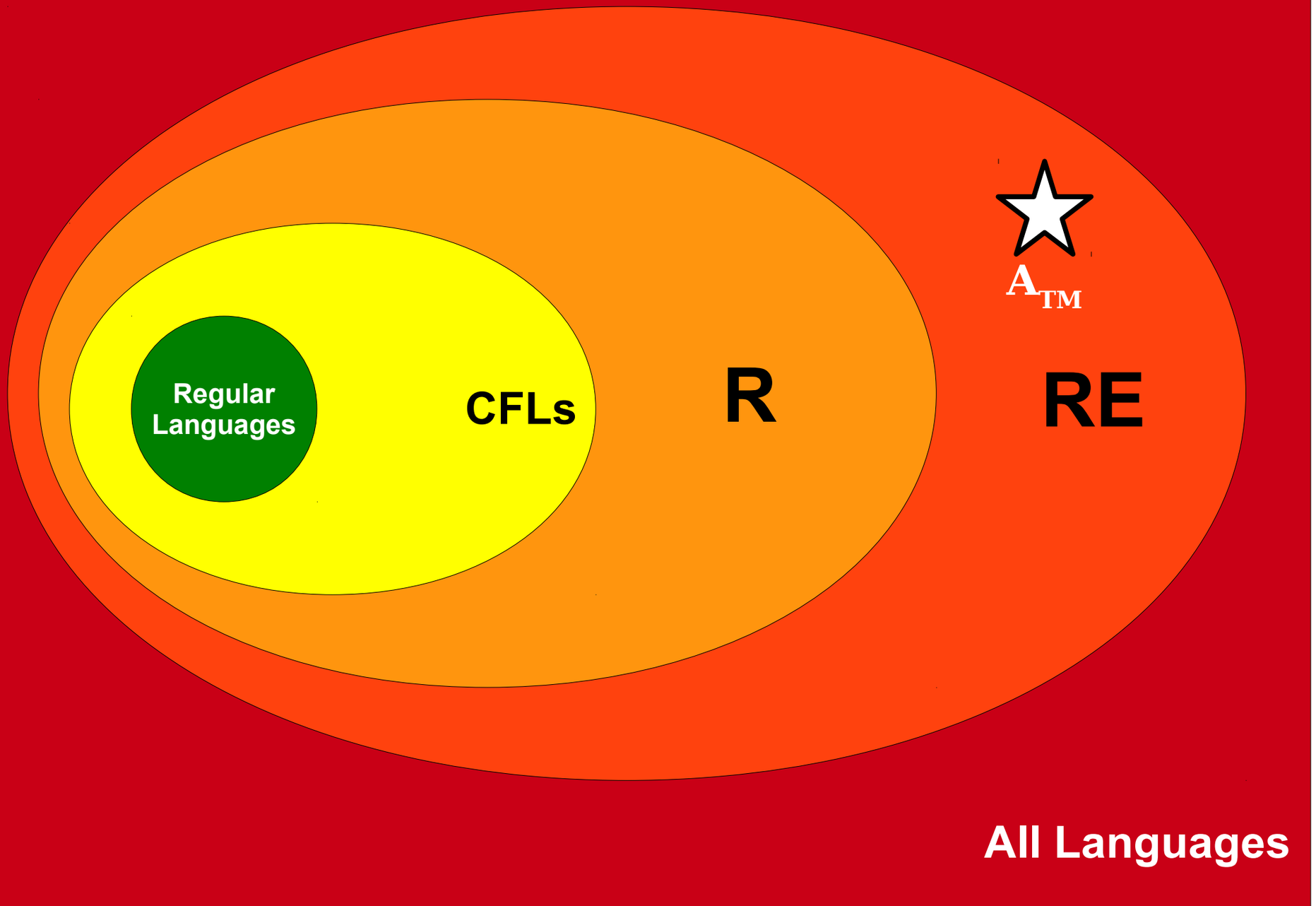
Given this, we could then construct this program P:

```
bool mystery(string input) {
    string me = mySource();

    if (willAccept(me, input)) return false;
    else return true;
}
```

Pick an arbitrary string w and trace through the execution of program P on input w. If `willAccept(me, input)` returns true, then P must accept its input w. However, in this case P proceeds to reject its input w. Otherwise, if `willAccept(me, input)` returns false, then P must not accept its input w. However, in this case P proceeds to accept its input w.

In both cases we reach a contradiction, so our assumption must have been wrong. Therefore, $A_{TM} \notin \mathbf{R}$. ■


Regular Languages
CFLs
R
RE
A_{TM}
All Languages

What Does This Mean?

- In one fell swoop, we've proven that
  - A_{TM} is ***undecidable***; there is no general algorithm that can determine whether a TM will accept a string.
  - $\mathbf{R} \neq \mathbf{RE}$, because $A_{TM} \notin \mathbf{R}$ but $A_{TM} \in \mathbf{RE}$.
- What do these two statements really mean? As in, why should you care?

$A_{TM} \notin \mathbf{R}$

- The proof we've done says that

  There is no possible way to design an algorithm that will determine whether a program will accept an input.

- Notice that our proof just assumed there was some decider for A_{TM} and didn't assume anything about how that decider worked. In other words, no matter how you try to implement a decider for A_{TM}, you can never succeed!

R ≠ RE

- Because **R ≠ RE**, there are some problems where “yes” answers can be checked, but there is no algorithm for deciding what the answer is.
- ***In some sense, it is fundamentally harder to solve a problem than it is to check an answer.***

More Undecidability Results

The Halting Problem

- The most famous undecidable problem is the ***halting problem***, which asks:

  Given a TM *M* and a string *w*,
  will *M* halt* when run on *w*?

- As a formal language, this problem would be expressed as

  ***HALT* = { ⟨*M*, *w*⟩ | *M* is a TM that halts on *w* }**

- How hard is this problem to solve?

* i.e., accept or reject, as opposed to infinite loop

$HALT \in \mathbf{RE}$

- ***Claim:*** $HALT \in \mathbf{RE}$.
- ***Idea:*** If you were certain that a TM M halted on a string w, could you convince me of that?
- Yes – just run M on w and see what happens!

```
bool checkHalt(TM M, string w) {
   feed w into M;
   while (true) {
      if (M is in an accepting state) accept();
      else if (M is in a rejecting state) accept();
      else simulate one more step of M running on w;
   }
}
```

HALT ∉ **R**

- ***Claim:*** *HALT* ∉ **R**.
- If *HALT* is decidable, we could write some function

```
bool willHalt(string program,
              string input)
```

  that accepts as input a program and a string input, then reports whether the program will halt when run on the given input.
- Then, we could do this...

What does this program do?

```
bool mystery(string input) {
   string me = mySource();
   if (willHalt(me, input)) { //decider
      while (true) {
         // loop infinitely
      }
   } else {
      accept();
   }
}
```

What does this program do?

```
bool mystery(string input) {
   string me = mySource();
   if (willHalt(me, input)) { //decider
      while (true) {
         // loop infinitely
      }
   } else {
      accept();
   }
}
```

Imagine running this program on some input. What happens if...

... this program halts on that input?
It loops on the input!

... this program loops on this input?
It halts on the input!

Theorem: *HALT* $\notin$ **R**.

Proof: By contradiction; assume that *HALT* $\in$ **R**. Then there's a decider *D* for *HALT*, which we can represent in software as a method `willHalt` that takes as input the source code of a program and an input, then returns true if the program halts on the input and false otherwise.

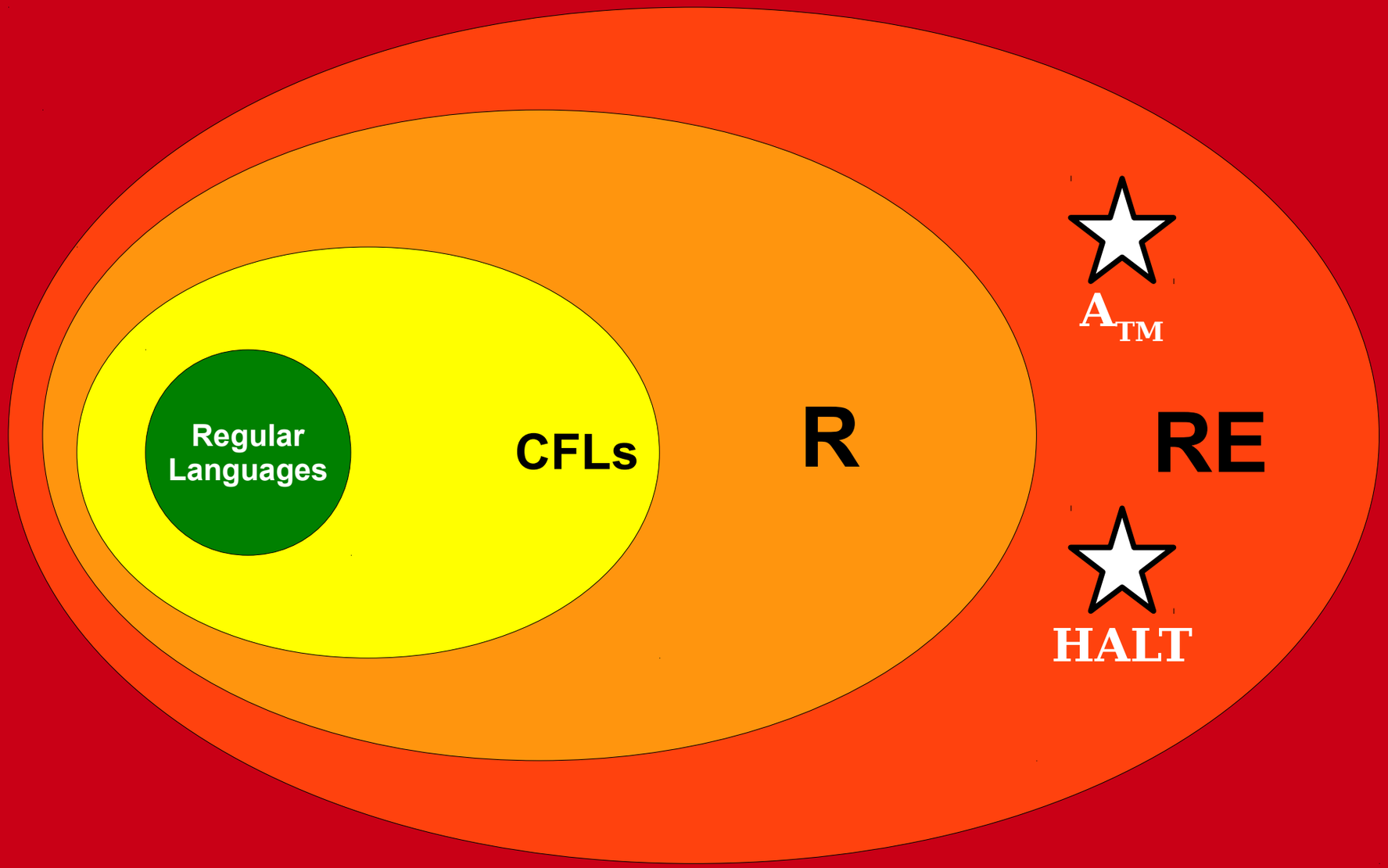
Given this, we could then construct this program *P*:

```
bool mystery(string input) {

   string me = mySource();
   if (willHalt(me, input)) { //decider
       while (true) {
           // loop infinitely
       }
   } else {
       accept();
   }
}
```

Choose any string *w* and trace through the execution of program *P* on input *w*, focusing on the answer given back by the `willHalt` method. If `willHalt(me, input)` returns true, then *P* must halt on its input *w*. However, in this case *P* proceeds to loop infinitely on *w*. Otherwise, if `willHalt(me, input)` returns false, then *P* must not halt its input *w*. However, in this case *P* proceeds to accept its input *w*.

In both cases we reach a contradiction, so our assumption must have been wrong. Therefore, *HALT* $\notin$ **R**. ■


Regular Languages
CFLs
R
RE
A_{TM}
HALT
All Languages