Theorem 0.0.1 (Cayley-Hamilton Theorem). *Suppose that A is an $n\times n$ matrix with characteristic polynomial $\chi_A(x)$. Then $\chi_A(A)$, the result of substituting the matrix A into the characteristic polynomial, is zero.*

Proof. We split into two cases.

- **Case 1: A has a T-invariant subspace.**

Suppose that $T : V \to V$ $(\dim V = n)$ is a linear transformation, and suppose that A has a T-invariant subspace that is not $\{\mathbf{0}_V\}$ or V. We proceed by induction on n in this case.

First, let $n = 1$. Then $A = \begin{bmatrix} a_{11} \end{bmatrix}$ and $\chi_A(x) = x - a_{11}$, so that $\chi_A(A) = \begin{bmatrix} a_{11} \end{bmatrix} - a_{11}I_1 = 0$.

Now assume that, for any $m < n$, the Cayley-Hamilton Theorem is true for the $m \times m$ matrix A.

Continue with the inductive step. Let the T-invariant subspace be W and let a basis of W be B_W. Let B be a basis of V produced by extending B_W. Obtain the basis $\bar{B}$ of the quotient space ${}^{V}\!/_{W}$ using the versions of the elements of $B \setminus B_W$ in the quotient space. Let $T|_W$ be the restriction of T to W and let $\bar{T}$ be the map corresponding to T in the quotient space. Recall that

$$[T]_B = \begin{bmatrix} [T|_W]_{B_W} & * \\ 0 & [\bar{T}]_{\bar{B}} \end{bmatrix}$$

(which is a block matrix). Note that

$$\chi_T(x) = \chi_{\bar{T}}(x)\chi_{T|_W}(x)$$

Using the property that if we apply a polynomial to a block matrix of the form of A then we merely need to apply it to each of the blocks inside the block matrix, we can write

$$\begin{aligned}
\chi_T(A) &= \chi_{T|_W}(A)\chi_{\bar{T}}(A) \\
&= \begin{bmatrix} \chi_{T|_W}\left([T|_W]_{B_W}\right) & * \\ 0 & \chi_{T|_W}\left([\bar{T}]_{\bar{B}}\right) \end{bmatrix} \begin{bmatrix} \chi_{\bar{T}}\left([T|_W]_{B_W}\right) & * \\ 0 & \chi_{\bar{T}}\left([\bar{T}]_{\bar{B}}\right) \end{bmatrix} \\
&= \begin{bmatrix} 0 & * \\ 0 & \chi_{T|_W}\left([\bar{T}]_{\bar{B}}\right) \end{bmatrix} \begin{bmatrix} \chi_{\bar{T}}\left([T|_W]_{B_W}\right) & * \\ 0 & 0 \end{bmatrix} \\
&= 0
\end{aligned}$$

where the evaluation of the characteristic polynomials came from the inductive hypothesis. This is what we wanted.

- **Case 2: A does not have a T-invariant subspace.**

For the other case, suppose that A has no non-trivial T-invariant subspaces for any $T : V \to V$. (Again we will let $\dim V = n$.) Define the conspicuously named B as $B = \{\mathbf{v}, T(\mathbf{v}), \dots, T^{n-1}(\mathbf{v})\}$ for whichever T you like and $\mathbf{v} \in V$ nonzero. We claim that B is a basis of V. To prove this, we will let j be the largest integer with

$$S_j = \{\mathbf{v}, T(\mathbf{v}), \dots, T^{j-1}(\mathbf{v})\}$$

a linearly independent set. This j must exist because clearly S_1 is linearly independent, also $1 \le j \le n$.

Now we must show $j = n$. Let $W = \operatorname{Span}(S_j)$, which means $\dim W = j$. Note that S_j is a basis of W. Now note that the set

$$S' = \{T(\mathbf{v}), T^2(\mathbf{v}), \dots, T^j(\mathbf{v})\}$$

is contained in W because, by definition of j, the set $S_j \cup \{T^j(\mathbf{v})\}$ is not linearly independent and therefore $T^j(\mathbf{v}) \in \operatorname{Span}(S_j) = W$. This means that $\mathbf{v} \in W \implies T(\mathbf{v}) \in W$, since S_j was a basis of W, and therefore W is T-invariant. Yet, we are told by our assumption that this requires W to be trivial. We can't have $W = \{\mathbf{0}_V\}$ since $j \ge 1$, so $j = n$, and B is a basis of V.

Now that we know this fact we can proceed. The matrix of T with respect to the basis B is

$$\begin{bmatrix} 0 & 0 & 0 & \dots & 0 & -a_0 \\ 1 & 0 & 0 & \dots & 0 & -a_1 \\ 0 & 1 & 0 & \dots & 0 & -a_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & -a_{n-1} \end{bmatrix}$$

where we let $T^n(\mathbf{v}) = -\left(a_0\mathbf{v} + a_1T(\mathbf{v}) + \dots + a_{n-1}T^{n-1}(\mathbf{v})\right)$. This matrix is exactly the companion matrix of the polynomial $\chi_T(x) = x^n + \sum_{i=0}^{n-1} a_i x^i$. With this observation, $\chi_T(T)(\mathbf{v})$ (which means the transformation $\chi_T(T)$ applied to the vector v) is $T^n(\mathbf{v}) + \left(a_{n-1}T^{n-1}(\mathbf{v}) + \dots + a_0\mathbf{v}\right)$, which is zero because of how we wrote out T^n.

But $\mathbf{v}$ was arbitrary, and $\chi_T(T)(\mathbf{v}) = \mathbf{0}_V$ for any $\mathbf{v} \in V$ implies that $\chi_T(T)$ is the zero linear map. $\square$